



Making the Most *of* Interim Assessment Data

Lessons from Philadelphia

June 2009





RESEARCH FOR ACTION

Research for Action (RFA) is a Philadelphia-based, non-profit organization engaged in policy and evaluation research on urban education. Founded in 1992, RFA seeks to improve the education opportunities and outcomes of urban youth by strengthening public schools and enriching the civic and community dialogue about public education. For more information about RFA please go to our website, www.researchforaction.org.

Learning from Philadelphia's School Reform

Research for Action (RFA) is leading *Learning from Philadelphia's School Reform*, a comprehensive, multi-year study of Philadelphia's school reform effort under state takeover. The project is supported with lead funding from the William Penn Foundation and related grants from Carnegie Corporation of New York, the Samuel S. Fels Fund, the Edward Hazen Foundation, the Charles Stewart Mott Foundation, The Pew Charitable Trusts, The Philadelphia Foundation, the Spencer Foundation, Surdna Foundation, and others.

Acknowledgements

We are deeply appreciative of the numerous overlapping communities of education researchers, practitioners, and activists of which we are a part. These communities sustain us and enrich our work. This research project, like so many others, has benefitted from these relationships. Below, we thank those who made specific contributions to this report.

The Spencer Foundation and the William Penn Foundation provided generous financial support for the research.

Researchers at the Consortium for Chicago School Research (CCSR) played an important role in the quantitative analysis. John Easton contributed to the research design and analysis and Steve Ponisciak, originally of CCSR and now at the Wisconsin Center for Education Research, conducted the analysis. We thank them for their technical expertise and their wisdom.

Many people were diligent readers and responders. The comments of two anonymous reviewers raised important questions that helped us to sharpen and cohere the report's content. Conversations and a joint project with colleagues at the Consortium for Policy Research in Education were helpful. Our colleagues at Research for Action – Diane Brown, Eva Gold, Tracey Hartmann, Rebecca Reumann, Elaine Simon and Betsey Useem – offered sage advice.

Getting a report to press is an arduous task. Judy Adamson, Managing Director at RFA, managed and directed the design, editing, and proofreading of the report. She was ably assisted by Joseph Kay, Philly Fellow extraordinaire, Judith Lamirand of Parallel Design, and Nancy Bouldin of Steege/Thomson Communications.

Most importantly, this report would not have been possible without the cooperation of the School District of Philadelphia. Staff in the Office of Accountability and Assessment provided the data that were needed and answered many questions. Central office administrators offered insights about the intentions of the district's Core Curriculum and Benchmark assessments. Staff of the district's Education Management Organization partners helped us gain access to schools. Special thanks to the principals, teacher leaders, and teachers in the ten schools in our qualitative sample. All gave graciously of their time, were patient with our many requests, and responded candidly to our questions. We are grateful to all of these people for all that they do for Philadelphia young people everyday.

Making the Most *of* Interim Assessment Data

Lessons from Philadelphia

Jolley Bruce Christman
Research for Action

Ruth Curran Neild
Johns Hopkins University

Katrina Bulkley
Montclair State University

Suzanne Blanc
Research for Action

Roseann Liu
University of Pennsylvania

Cecily Mitchell
Research for Action

Eva Travers
Swarthmore College

*The Consortium for Chicago School Research provided
technical assistance for the statistical analyses.*



RESEARCH FOR ACTION

Copyright © 2009 Research for Action

A report from *Learning from Philadelphia's School Reform*

The School District of Philadelphia

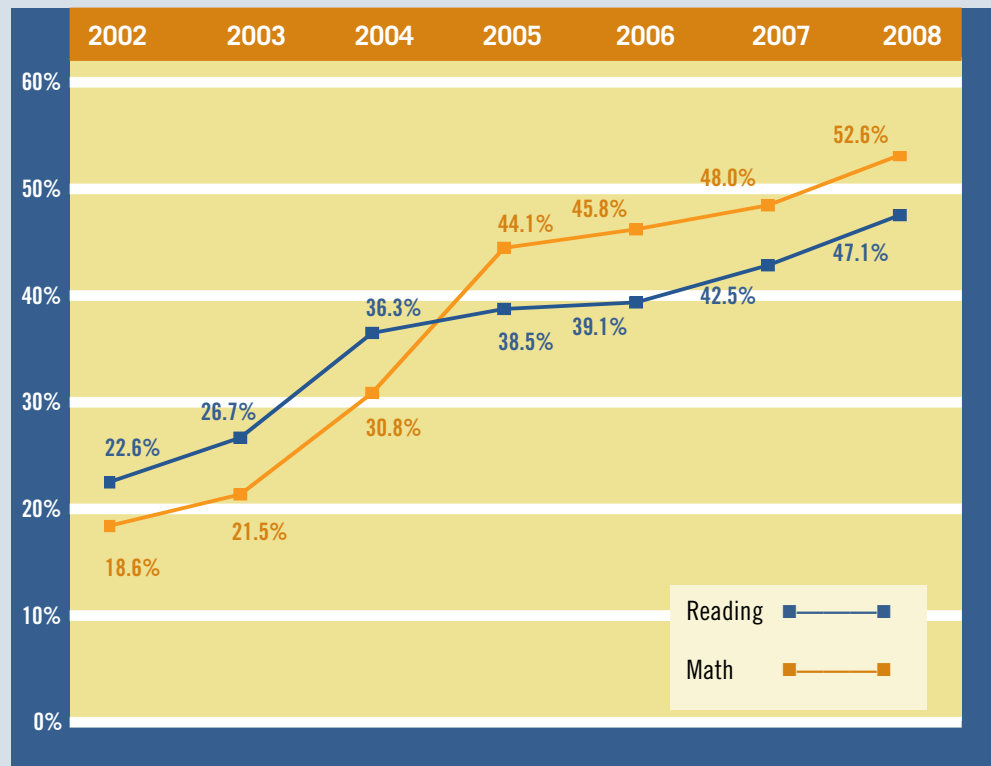
The School District of Philadelphia is the eighth largest district in the nation. In 2006-07 it enrolled 167,128 students. 62.4% of the students were African American, 16.9% were Latino, 13.3% were Caucasian, 6.0% were Asian, 0.2% were Native American, and 1.2% classified as Other.

In December 2001, the Commonwealth of Pennsylvania took over the School District of Philadelphia, declaring the city's schools to be in a state of academic and fiscal crisis, disbanding the school board and putting in place a School Reform Commission. In 2002, Paul Vallas became the CEO of the School District of Philadelphia. During his time as CEO from 2002 to 2007, student achievement scores rose substantially. The percentage of fifth and

Figure A.1 School District of Philadelphia 2002-2008 PSSA Results

Percentage of Students Advanced or Proficient, Grades 3-8 Combined

Initially grades 5 & 8. Grade 3 added in 2006, grades 4, 6, 7 added in 2007.



eighth graders (the grades consistently tested) scoring “Proficient” or “Advanced” on the Pennsylvania System of School Assessment (PSSA) tests went up 26 percentage points in math. In reading, the percentage went up by 11 points in fifth grade and 25 points in eighth grade. The percentage scoring in the lowest category (Below Basic) dropped in all tested grades by 26 points in math and 12 points in reading.

Test scores continued their climb in the year following Vallas's resignation when the district was led by an interim CEO who continued the same reforms. Achievement gains occurred despite serious under-funding by the state (Augenblick, Palaich and Associates, Inc., 2007) and despite the city's high and growing rate of poverty, the highest among the nation's 10 largest cities (Tatian, Kingsley, and Hendey, 2007).

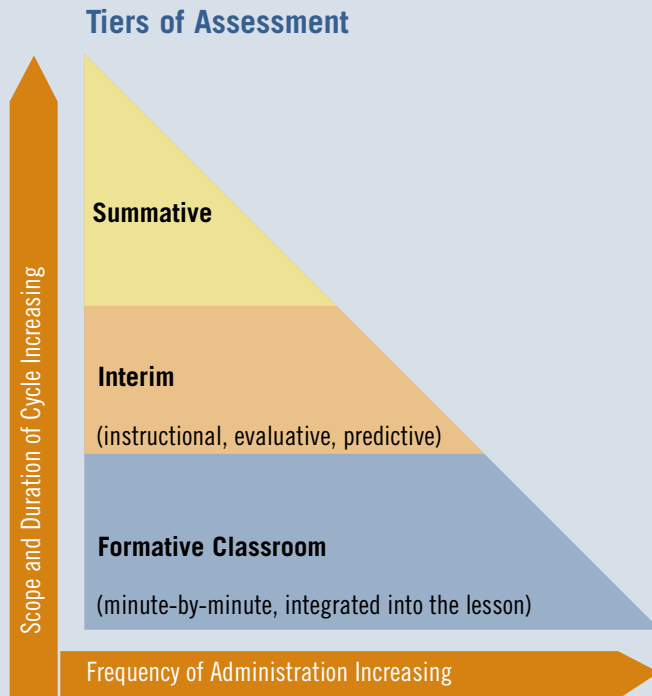
Table of Contents

| | |
|--|-----------|
| Introduction | 1 |
| The Usefulness of Interim Assessments: Competing Claims | 2 |
| Overview of Report | 3 |
| Chapter 1 - Organizational Learning: A Framework for Examining the Use of Benchmark Assessment Data | 5 |
| Conceptual Framework | 7 |
| Research Methodology | 8 |
| Research Questions | 9 |
| Chapter 2 - Philadelphia's Managed Instruction System | 15 |
| The Philadelphia Context | 15 |
| The Core Curriculum | 17 |
| SchoolNet | 20 |
| Benchmark Assessments | 21 |
| In Summary | 27 |
| Chapter 3 - The Impact of Benchmarks on Student Achievement | 31 |
| The Organizational Learning Framework and Key Research Questions | 31 |
| Analytic Approach | 33 |
| Findings | 38 |
| In Summary | 44 |
| Chapter 4 - Making Sense of Benchmark Data | 45 |
| Three Kinds of Sense-Making: Strategic, Affective, and Reflective | 46 |
| Making Sense of Benchmarks: Four Examples | 48 |
| In Summary | 53 |
| Chapter 5 - Making the Most of Benchmark Data: The Case of Mahoney Elementary /School | 57 |
| School Leaders and Effective Feedback Systems | 59 |
| Grade Group Meetings and Benchmark Discussions | 61 |
| Organizational Learning and Instructional Coherence | 63 |
| Conclusion - Making the Most of Interim Assessment Data: Implications for Philadelphia and Beyond | 65 |
| Investing in School Leaders | 65 |
| Designing Interim Assessments and Supports for their Use | 67 |
| Implications for Further Research | 68 |
| References | 69 |
| Appendices | 72 |
| Authors | 82 |

Three Kinds of Assessments

Perie, Marion, Gong, and Wurtzel (2007)¹ have categorized the three kinds of assessments currently in use — summative, formative, and interim — by their intended purposes, audiences, and the frequency of their administration.

- **Summative** assessments are given at the end of a semester or year to measure students' performance against district or state content standards. These standardized assessments are often part of an accountability system and are not designed to provide teachers with timely information about their current students' learning.
- **Formative** assessments occur in the natural course of teaching and learning. They are built into classroom instructional activities and provide teachers and students with ongoing, daily information about what students are learning and how teachers might improve instruction so that learning gaps and misunderstandings can be remedied. These assessments do not provide information that can be aggregated.
- **Interim** assessments fall between formative and summative assessments and provide standardized data that can be aggregated. Interim assessments vary in their purpose. They may predict student performance on an end-of-year summative, accountability assessment; they may provide evaluative information about the impact of a curriculum or a program; or, they may offer instructional information that helps diagnose student strengths and weaknesses.



Source: Perie et al. (2007)

Figure A.2

¹ Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007, November). *The role of interim assessments in a comprehensive assessment system*. Washington, DC: The Aspen Institute.

Introduction

In recent years, school reformers have embraced data-driven decision-making as a central strategy for improving much of what is wrong with public education. The appeal of making education decisions based on hard data – rather than tradition, intuition, or guesswork – stems partly from the idea that data can make the source of a problem clearer and more specific. This newfound clarity can then be translated into sounder decisions about instruction, school organization, and deployment of resources.

In urban districts, the press for data-driven decision-making has intensified in the stringent accountability environment of No Child Left Behind, where schools look for ways to increase their students' performance on state assessments. These districts increasingly are turning to the significant for-profit industry that has sprung up to sell them curricula aligned with state standards, data management systems, and interim assessments.² Interim assessments are standardized assessments administered at regular intervals during the school year in order for educators to gauge student achievement before the annual state exams used to measure Adequate Yearly Progress (AYP). Results of interim assessments can be aggregated and reported at a variety of levels, usually classroom, grade, school and district. The tools for administering and scoring the assessments and storing, analyzing, and interpreting the assessment data are being marketed by vendors as indispensable aids to meeting NCLB requirements.³

In this report, Research for Action (RFA) examines the use and impact of interim assessment data in elementary schools in the School District of Philadelphia. Philadelphia was an early adopter of these assessments, implementing them district-wide in September 2003. The report presents findings from one of the first large-scale empirical studies on the use of interim assessments and their impact on student achievement.

Interim assessments are a central component of what the School District of Philadelphia's leaders dubbed a "Managed Instruction System" (MIS). The MIS includes a Core Curriculum and what are called Benchmarks in Philadelphia. Benchmark assessments were developed in collaboration with Princeton Review, a for-profit company, and are aligned with the Core Curriculum. In Philadelphia, classroom instruction in grades three through eight occurs in six-week cycles: five weeks of instruction, followed by the administration of Benchmark assessments. In one or two days between the fifth and sixth weeks, teachers analyze Benchmark data and develop instructional responses to be implemented in the sixth week.

The Philadelphia Benchmarks are consistent with the definition of interim assessments offered by Perie, Marion, Gong and Wurtzel (2007) in that the

² Burch, P. (2005, December 15). The new education privatization: Educational contracting and high stakes accountability. *Teachers College Record*.

³ Burch, P. (2005, December 15).

Benchmarks: “(1) assess students’ knowledge and skills relative to curriculum goals within a limited time frame, and (2) are designed to inform teachers’ instructional decisions as well as decisions beyond the classroom levels.”⁴ (See [Figure A-2](#) for a description of the differences among three kinds of assessments — summative, interim, and formative assessments.)

The Usefulness of Interim Assessments: Competing Claims

The introduction of interim assessments in urban districts across the country has not been without controversy, as district leaders, teachers, and the testing industry make conflicting claims for the efficacy of these assessments for guiding instruction and improving student achievement. Many educators and assessment experts, alarmed by the growing market in off-the-shelf commercial products labeled as “formative” assessments, insist that the only true formative assessments “must blend seamlessly into classroom instruction itself.”⁵ There is good evidence that these instructionally embedded assessments have a positive effect on student learning.⁶ In theory, at least, interim assessments could be expected to have a similarly beneficial effect on teaching and learning as instructionally embedded, “formative” classroom assessments. To date, however, there is not the same kind of empirical base for the claim that interim assessments have the power of classroom-based assessments. And, for a number of reasons, it can not be assumed that they would have the same positive impact. For example, because interim assessments do not occur at the time of instruction, they may not provide the kind of immediate feedback that is useful to teachers and students. And because they are standardized tests that almost always rely on a multiple choice format, they may not offer adequate information about “*how* students understand.”⁷

The controversy over interim assessments is growing as district budgets shrink and there remains little empirical evidence about the efficacy of the assessments in improving student achievement. The Providence Public School District abandoned its quarterly assessments after three years of implementation. Researchers who documented Providence’s experience noted, “District-level administrators provided a variety of explanations for the decision, including a lack of evidence of effectiveness and the summative character of the assessments, but left open the possibility of reinstating the assessments at a

Unraveling the benefits of interim assessment data to improvement in student learning is a necessarily complex task.

⁴ Perie, M. et al., 2007, p. 4.

⁵ Cech, S. J. (2008, September 17). Test industry split over ‘formative’ assessments. *Education Week*, 28(4), 1, 15, p. 1.

⁶ Black, P. & William, D. (1998, October). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*.

⁷ Perie, M. et al., 2007, p. 22.

later date.”⁸ In January 2009, the Los Angeles teachers union threatened to boycott the “periodic assessments” mandated by the district – a series of exams given three or four times a year at secondary schools – claiming that the tests are costly and counterproductive. Such district tests at all grade levels “have become central to a debate over the proliferation of testing, whether it interrupts instruction and can narrow the depth and breadth of what’s taught.”⁹

Overview of Report

Our research shows that Philadelphia’s elementary school teachers – in contrast to those in some other districts, such as Los Angeles, – have embraced the Benchmark assessments, finding them useful guides to their classroom instruction. However, unraveling the benefits of the Benchmark data to improvement in student learning is a necessarily complex task. In this study, we use data from a district-wide teacher survey, student-level demographic and achievement data, and qualitative data obtained from field observations and interviews to examine the associations among such factors as instructional leadership, a positive professional climate among teachers, teacher investment in the Core Curriculum and Benchmarks, and gains in student achievement on standardized tests.

Our analysis indicates that teachers’ high degree of satisfaction with the information that Benchmark data provide is not itself a statistically significant predictor of student achievement gains. However, used in tandem, the Core Curriculum and Benchmarks have established clear expectations for what teachers should teach and at what pace. And, importantly, students in schools where teachers made more extensive use of the Core Curriculum made greater achievement gains than in schools where teachers used it less extensively.

Benchmarks’ alignment with the Core Curriculum offers the opportunity for practitioners to delve more deeply into the curriculum as they review Benchmark results, thereby reinforcing and strengthening use of the curriculum. Surprisingly, however, our qualitative research showed that Philadelphia’s school leaders and teachers are not capitalizing on Benchmark data to generate deep discussions of and learning about the Core Curriculum. This suggests that continued use of Benchmark assessments in Philadelphia is not likely to contribute to improved student learning without greater attention to developing strong principals and teacher leaders. These school leaders need to know how to facilitate probing conversations that promote teachers’ learning

⁸ Clune, W. H. & White, P. A. (2008, October). Policy effectiveness of interim assessments in Providence Public Schools. WCER Working Paper No. 2008 Wisconsin Center for Education Research, School of Education, University of Wisconsin-Madison <http://www.wcer.wisc.edu/>. p. 5.

⁹ Blume, H. (2009, January 28). L.A. teachers' union calls for boycott of testing. Los Angeles Times [On-line]. Retrieved on February 11, 2009 from <http://www.latimes.com/news/education/la-me-laUSD28-2009jan28,0,4533508.story>.

about curriculum and pedagogy. In this report, we use an organizational learning framework to offer specific recommendations for what district leaders can do to help school staff make the most of Benchmark results.

It is important to note that while our research reviews how Philadelphia deployed its assessment model and examines student achievement data to assess its impact, this report should not be seen as a review of the technical quality of Philadelphia's Benchmark assessments, interim assessments in general, or the Core Curriculum. A close examination of the technical merits of these elements of the managed instruction system was beyond the scope of this project.¹⁰

Chapter One outlines our conceptual framework for interim assessments and organizational learning, identifies key research questions, and summarizes the research methodology of this study.

In Chapter Two, we describe Philadelphia's Managed Instruction System, highlighting district leaders' expectations for how school staff would use its components. We draw on data from the district-wide teacher survey to describe teachers' use of the Core Curriculum and satisfaction with the Benchmark assessments.

In Chapter Three, we address the question of whether the Managed Instruction System and supportive school conditions for data use were associated with greater student learning gains.

Chapter Four describes how school staff make sense of Benchmark data and consider their implications for instruction. What do school leaders and teachers talk about and what plans do they make as a result of their interpretation of the data?

Chapter Five is a case study of the Mahoney Elementary School. This case provides concrete images of what school leaders and instructional communities can do to enrich the use of Benchmark data.

In the Conclusion, we discuss implications of this research for what needs to be done in order for school staff to make the most of interim assessment data.

Philadelphia's elementary school teachers have embraced the Benchmark assessments, finding them useful guides to their classroom instruction.

¹⁰ In 2005, Phi Delta Kappa International issued its assessment of the Core Curriculum and the Benchmark assessments in "A Curriculum Management Audit in Literacy and Mathematics of the School District of Philadelphia." The report has only recently become available. Its authors found that while the Core Curriculum had provided consistence in what is taught, 87 percent of its instructional strategies in mathematics are at the knowledge and comprehension levels. When the auditors observed classroom instruction, they found that 84 percent of the instructional strategies used were at the knowledge and comprehension levels. Their overall judgment was that the School District of Philadelphia was not meeting its own expectations for a rigorous curriculum. In reviewing the Benchmark assessments, they also judged that most of the items composing the test were at the levels of knowledge and comprehension.

Chapter One

Organizational Learning: A framework for examining the use of Benchmark assessment data

Teaching is a complex enterprise. In order to help each student learn, a teacher must be aware of the needs and strengths of individual students and the class as a whole. She must note how children are making sense of newly introduced concepts and how they are developing increasingly advanced skills. What have children mastered and what continues to pose difficulty for them? What is helping them learn? What is getting in their way?

The logic behind how interim Benchmark assessment data can assist teachers is straightforward: a teacher acquires data about what her students have learned; she examines the data to see where her students are strong and weak; she custom-tailors what and how she teaches so that individuals and groups of students learn more; and as teachers across the school engage in this process, the school as a whole improves.

While we recognize the importance of an individual teacher's use of student performance data to guide her instruction, this report views use of student data through a different lens. Specifically, we explore how an **organizational learning framework** can inform our understanding of how to strengthen the capacity of schools to capitalize on Benchmark and other kinds of data.

Our focus on organizational learning follows from the school change literature which indicates that in order for all students to make consistent academic progress, school staff must work together in **concerted** ways to advance the quality of the educational program.¹¹ School improvement is a problem of organizational learning, that is, the ability of school leaders and teachers to identify and problem-solve around constantly changing challenges. From the perspective of organizational learning, urban schools – like other organizations – will be better equipped to meet existing and future challenges “by creating new ways of working and developing the new capabilities needed for that work.”¹²

¹¹ Little, J. W. (1999). Teachers' professional development in the context of high school reform: Findings from a three-year study of restructuring high schools. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Quebec.; Wagner, T. (1998). Change as Collaborative Inquiry: A 'Constructivist' Methodology for Reinventing Schools. *Phi Delta Kappan*, 80(7), 378-383.; Knapp, M. S. (1997). Between Systemic Reforms and the Mathematics and Science Classroom: The Dynamics of Innovation, Implementation, and Professional Learning. *Review of Educational Research*, 67(2), 227-266.; Spillane, J. P. & Thompson, C. L. (1997, June). Reconstructing Conceptions of Local Capacity: The Local Education Agency's Capacity for Ambitious Instructional Reform. *Education Evaluation and Policy Analysis*, 19(2), 185-203.; Senge, P. (1990). *The Fifth Discipline: The Art & Practice of the Learning Organization*. NY: Doubleday.

¹² Resnick, L. B. & Hall, M. W. (1998). Learning Organizations for Sustainable Education Reform. *Journal of the American Academy of Arts and Sciences*, 127(4), 89-118, p. 108.

Recent research has begun to address the multiple factors related to **overall organizational capacity** that affect data use.¹³ School capacity incorporates multiple aspects of schools and the literature suggests that school capacity has four dimensions:

- human capital (the knowledge, dispositions, and skills of individual actors);
- social capital (social relationships characterized by trust and collective responsibility for improved organizational outcomes);
- material resources (the financial and technological assets of the organization);¹⁴ and
- structural capacity (an organization's policies, procedures, and formal practices).¹⁵

An important feature of learning organizations is the existence of a relational culture that is characterized by collaboration, openness, and inquiry.¹⁶ Knowledge building is a collective process that involves the development of a shared language and commonly held beliefs. Organizational knowledge “is most easily generated when people work together in tightly knit groups.”¹⁷ Applying this theory, we examined how formal instructional communities made sense of data from Benchmark assessments and generated actionable knowledge for planning instructional improvements.

A second focus of the study, also drawn from organizational learning theory is the use of student performance data within feedback systems composed of “structures, people, and practices” that help practitioners transform data into actionable knowledge.¹⁸ In our effort to understand how Benchmark data contribute to organizational learning, we applied the concept of a four-step “feedback system” to analyze the structures and processes educators use to engage with data collectively and systematically during the course of a

An important feature of learning organizations is the existence of a relational culture that is characterized by collaboration, openness, and inquiry.

¹³ Mason, S. A. & Watson, J. G. (2003). Understanding Schools' Capacity to Use Data. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL; Leithwood, K., Aitken, R., & Jantzi, D. (2001). *Making Schools Smarter: A System for Monitoring School and District Progress*. Thousand Oaks, CA: Corwin Press.

¹⁴ Spillane, J. P. & Thompson, C. L., 1997.

¹⁵ Century, J. R. (2000). Capacity. In N. L. Webb, J. R. Century, N. Davila, D. Heck, & E. Osthoff (Eds.), *Evaluation of systemic change in mathematics and science education*. Unpublished manuscript, University of Wisconsin-Madison, Wisconsin Center for Education Research.

¹⁶ Senge, P., 1990; Argyris, C. & Schon, D. A. (1978). *Organizational learning: A theory of action perspective*. Reading, MA: Addison-Wesley.

¹⁷ Brown, J. S. & Duguid, P. (1998). Organizing knowledge. *California Management Review*, 40(3), 28-44, p. 28.

¹⁸ Halverson, R. R., Prichett, R. B., & Watson, J. G. (2007). Formative feedback systems and the new instructional leadership (WCER Working Paper No. 2007-3). [On-line]. Retrieved on July 16, 7 A.D., from <http://www.wcer.wisc.edu/publications/workingPapers/index.php>.

school year. The four steps in the feedback system are; 1) accessing and organizing data, 2) sense-making to identify problems and solutions, 3) trying solutions, and 4) assessing and modifying solutions.

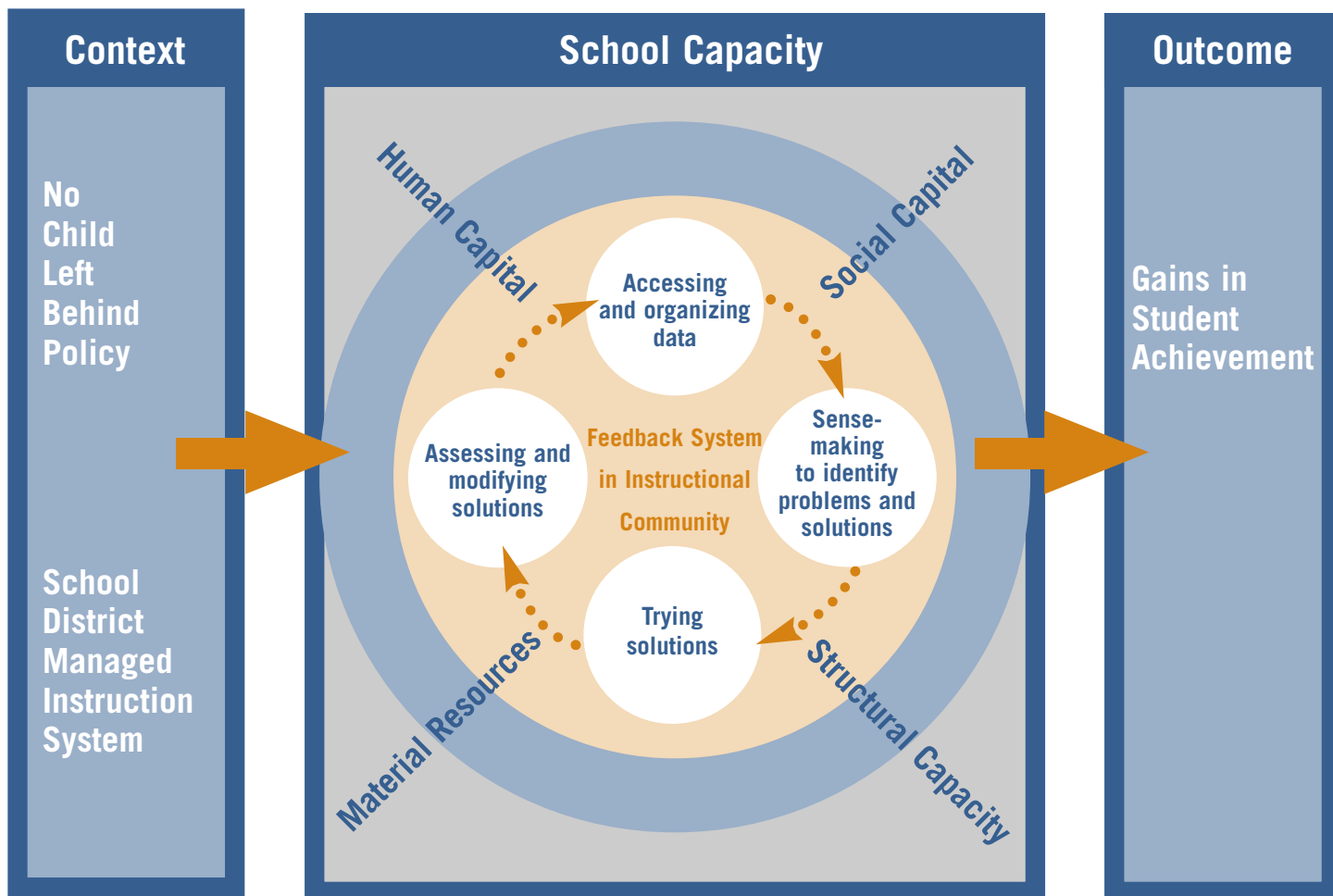
Conceptual Framework

The conceptual framework that guided our research, illustrated in [Figure 1.1](#), reflects the ideas discussed above. On the left, the figure depicts the larger policy and management context that we hypothesize will influence use of Benchmark data – the school district’s Managed Instruction System and the larger accountability environment of No Child Left Behind (NCLB). The middle box represents the four dimensions of school capacity discussed above. In this study, we focus on the role of school leaders and instructional communities in strengthening school capacity. An organizational framework suggests that these actors will be critical for creating the organizational practices necessary for coherent feedback systems that strengthen organizational learning and school improvement. The four-step feedback system described above is embedded within overall school capacity and instructional communities. It is important to note that multiple feedback systems will be operating simultaneously in a school; that these feedback systems do not operate in a lock-step manner and are most likely to be iterative; and, that, in the ideal, knowledge generated from one feedback system will inform other feedback systems. Finally, on the right, we anticipate that the outcome of these processes will be reflected in gains in student achievement.

This model highlights the complexity of data-driven decision-making and the use of Benchmark data to guide instruction. For example, it implies that if any one of the links in the feedback system in instructional communities is missing – that is, if teachers do not examine student data or do not know how to interpret the data they receive, or if they do not make instructional decisions that follow logically from a careful interpretation of the data, or if these decisions are not actually implemented in the classroom, or if their effectiveness is not assessed – the potential to increase student achievement is weakened. Further, it implies that the relative skill with which each activity is carried out – for example, whether the instructional decisions that arise from the data are excellent or merely adequate – can affect how much students learn.

The model also highlights the human, social, and material conditions in the school that increase the likelihood of teachers being able to make good use of student data. For example, strong school leadership is hypothesized to have a positive effect on teachers’ opportunities to access and interpret data and make appropriate instructional adjustments. School leadership also will affect

Figure 1.1 Conceptual Framework



the extent to which teachers are encouraged to use elements of a Managed Instruction System, including the Core Curriculum. In addition, the material conditions of the school, including access to computers and the Internet, may affect the extent to which teachers are able to review student data.

Research Methodology

This study includes information from the period September 2004 through June 2007. During the first year of the project, the research was exploratory in nature and focused on learning about the district's Managed Instruction System as it unfolded, identifying schools that exemplified effective use of data, and working with the district to develop and pilot a district-wide teacher survey that included items related to data use. The report draws on three kinds of data:

- a district-wide teacher survey administered in the spring of 2006 and 2007;
- student-level demographic and achievement data from standardized tests; and
- qualitative data obtained from intensive fieldwork in ten elementary schools and interviews with district staff and others who worked with the schools, as well as further in-depth case study analysis of five schools in 2006-2007.

Teacher Survey Data

The district's Office of Accountability and Assessment constructed a single teacher survey that combined questions about different topics. From the perspective of this study, important survey items included questions about school leadership, climate, and collegiality, developed and documented by the Consortium on Chicago School Research. The survey also included several original questions specific to Benchmarks, such as satisfaction with Benchmarks, professional development on data use, access to technology that could enable viewing student data online, and discussion of instructional responses to data with fellow teachers and school leaders. While these data-related survey questions provide important insights, a more complete understanding of the use of, and professional development for, Benchmarks and other types of student data would have required a considerably longer set of items. However, we use what is available to us to identify associations between data-related variables, school leadership and climate, and student achievement. In addition, teachers were asked about the subject(s) they taught and the grade span in which they were teaching. (NOTE: In Chapters Two and Three, we provide more information about the district-wide teacher survey, the sample for our study, and our analytic approach.)

Student Test Score Data

Our analysis relies on measurement of student academic growth obtained from longitudinal data on student achievement made available by the School District of Philadelphia. Student test score data from spring 2005, 2006, and

Research Questions

- 1 What were district leaders' expectations for how school staff would use Benchmark data and what supports did they provide to help practitioners become proficient in using data to guide instruction?
- 2 Were teachers responsive to the Managed Instruction System, particularly the Benchmark assessments? Did they use them? Did they find them helpful?
- 3 Did students experience greater learning gains at schools where the conditions were supportive of data use: that is, where the Managed Instruction System was more widely accepted and used and where analysis of student data was more extensive?
- 4 What can school leaders do to ensure that the use of Benchmark data contributes to organizational learning and ongoing instructional improvement within and across instructional communities?

2007 were analyzed for students who were in grades 4 through 8 during 2005-2006 and/or 2006-2007. The tests were either the Terra Nova or assessments from the PSSAs, depending on the grade and year. Raw scores for each student were converted to their percentile score within the district during the year and these scores then were converted to z-scores with a mean of zero and a standard deviation of one. To create a measure of growth, we examine changes in students' performance on standardized tests given at the end of successive school years. This strategy examines the "value added" to learning by attending a school in a given year. In this report, we examine improvement in student academic growth in two school years (2005-2006 and 2006-2007) for students in 4th through 8th grades.

Qualitative Data

The goal of our school-based qualitative research and in-depth case study research was to develop a fine-grained analysis of the dynamic interactions among school leadership, data use by instructional communities (grade groups), and instructional planning. Our aim was to identify the micro-practices of school leaders and instructional communities as they worked with data and put into action the resulting instructional decisions. Micro-practices refer to the routine actions that are part of the larger function of data-driven decision-making. Examples of micro-practices include: how data are formatted for analysis; how leaders facilitate discussions of data among staff; and, how they communicate messages about the importance of data.

The school sample was composed of ten elementary schools that were among the 86 schools identified as "low performing" and eligible for intervention under a state takeover of the School District of Philadelphia. The 86 low-performing schools represented 39 percent of the district's 220 elementary and middle schools. Like the other 76 low-performing schools, each of the ten schools in our sample was assigned to an intervention model beginning in the 2002-2003 school year. Seven of the schools were under management by outside providers; two schools were part of the district's homegrown intervention under the Office of Restructured Schools; one school was a "sweet sixteen" school – a low-performing school that was showing improvement and therefore received additional resources for two years but was not under a special management arrangement. We chose to take an in-depth look at the use of Benchmark data in low-performing schools because these schools were under considerable pressure to improve test scores and they had more resources, including, in most cases, additional personnel to provide support

for data use. We believed that these two factors would increase the likelihood that they would turn to the Benchmark data for guidance.¹⁹

In identifying schools to be part of the qualitative study, we sought out schools from each intervention model that would provide insight about how schools learn to engage with data as part of a process of school change. We developed a purposive sample of schools that were identified by district staff, provider staff, and other school observers as being well on the road to making effective use of data. Criteria for selection included: data-driven decision-making was a stated priority of school leaders; professional development on how to access, organize and interpret Benchmark data was ongoing; and, grade group meetings to discuss Benchmark data occurred regularly.

All of our schools served a considerably higher percentage of students living in poverty than the district average and served student populations that were predominantly either African American or Latino. (See Appendix A for more information about the ten schools.) It should be noted that, during the course of our study, the majority of these 10 schools were undergoing organizational restructuring. CEO Vallas believed that K-8 schools were more hospitable environments for middle grades students and either closed or converted most of Philadelphia's middle schools into K-8 schools and added grades 6-8 to many elementary schools.

In 2005-2006, a team of at least two researchers made two one-day site visits to each of the ten schools. During the visit, we conducted semi-structured interviews with the principal and two or three teacher leaders. Interviews lasted 60-90 minutes. (See Appendix C for lists of topics covered in the interviews.) Site visits were scheduled on days when we also could observe a leadership team meeting, grade group meeting(s), or other data related event(s).

In 2006-07, we narrowed our sample to five schools for more intensive fieldwork. To select the five schools, we developed criteria related to four categories; the principal's role in data use, the strength of the professional community, the school's AYP status, and the purposes that school staff brought to their interpretation of Benchmark data. The research team placed schools along continua for each category and selected schools that represented the range of variation. Two researchers spent about four days in each school. During these visits, we followed up with principals, observed several events at which staff discussed data, talked extensively with teacher leaders, and also interviewed at least two classroom teachers in each school. By June

¹⁹ In addition, an original intention of the study was to use the different management models as points of comparison. However, this research purpose fell away when all of the provider organizations, except Edison Schools, Inc. adopted the district's Managed Instruction System.

Table 1.1 School-Based Interviews and Observations

| Type of Respondent | 2004-05 | 2005-06 | 2006-07 | Total |
|---|-----------|-----------|-----------|-------|
| Principal | 6 | 17 | 9 | 32 |
| Subject Area Teacher Leader | 13 | 24 | 13 | 50 |
| Teacher | 5 | 23 | 28 | 56 |
| Other School Leader (e.g., Ass't Principal) | 1 | 3 | 12 | 16 |
| Total # of Interviews Conducted | 25 | 67 | 62 | |
| Type of Observation | 2004-05 | 2005-06 | 2006-07 | Total |
| Grade Group Meeting | 2 | 8 | 4 | 14 |
| Leadership Team Meeting | 0 | 5 | 5 | 10 |
| Professional Development Session | 10 | 3 | 6 | 19 |
| Other Event (e.g., CSAP meeting) | 0 | 8 | 3 | 11 |
| Total # of Observations Conducted | 12 | 24 | 18 | |

2007, our qualitative data set included more than 150 interviews with school staff and faculty; 54 observations of leadership team meetings, grade group meetings, and school-wide professional development sessions; and a collection of school documents. (See [Table 1.1](#))

RFA's qualitative research also included six interviews with administrators from the district's offices of Accountability, Assessment, and Intervention; Curriculum; and Professional Development. The topics covered included the Core Curriculum; student performance assessments generally, as well as in-depth probing about Benchmark assessments; professional development for school leaders on using data; and perceptions of whether and how the different providers operating in the district were using the district's Core Curriculum and Benchmark system. Researchers also interviewed staff from the education provider organizations to understand the policies and supports related to data use offered by these organizations to the schools that they were managing. (See [Table 1.2](#))

To analyze the interviews, we coded the data using a software package for qualitative data analysis and identified themes and practices within and across schools and providers using content analysis. We used information from written documents and field observations to triangulate our findings.

Table 1.2 Central Office and Provider Interviews

| Interviewee | 2004-05 | 2005-06 | 2006-07 | Total |
|----------------|---------|---------|---------|-------|
| Central Office | 2 | 4 | 0 | 6 |
| Provider | 9 | 2 | 0 | 11 |
| Total | 11 | 6 | 0 | |

Other analytical strategies included: case study write-ups of data use in each of the ten schools; reduction of data into word charts (for example, a chart describing the types of data that were attended to by school staff, the settings and actors involved, and the resulting instructional decisions); and development of extended vignettes of feedback systems in schools. More specific details on research methods, data analysis, and sample instruments can be found in Appendices B, C, D, and E.

In the next chapter, we take a closer look at the design of the Managed Instruction System and district leaders' expectations for use of the Benchmark assessment data.

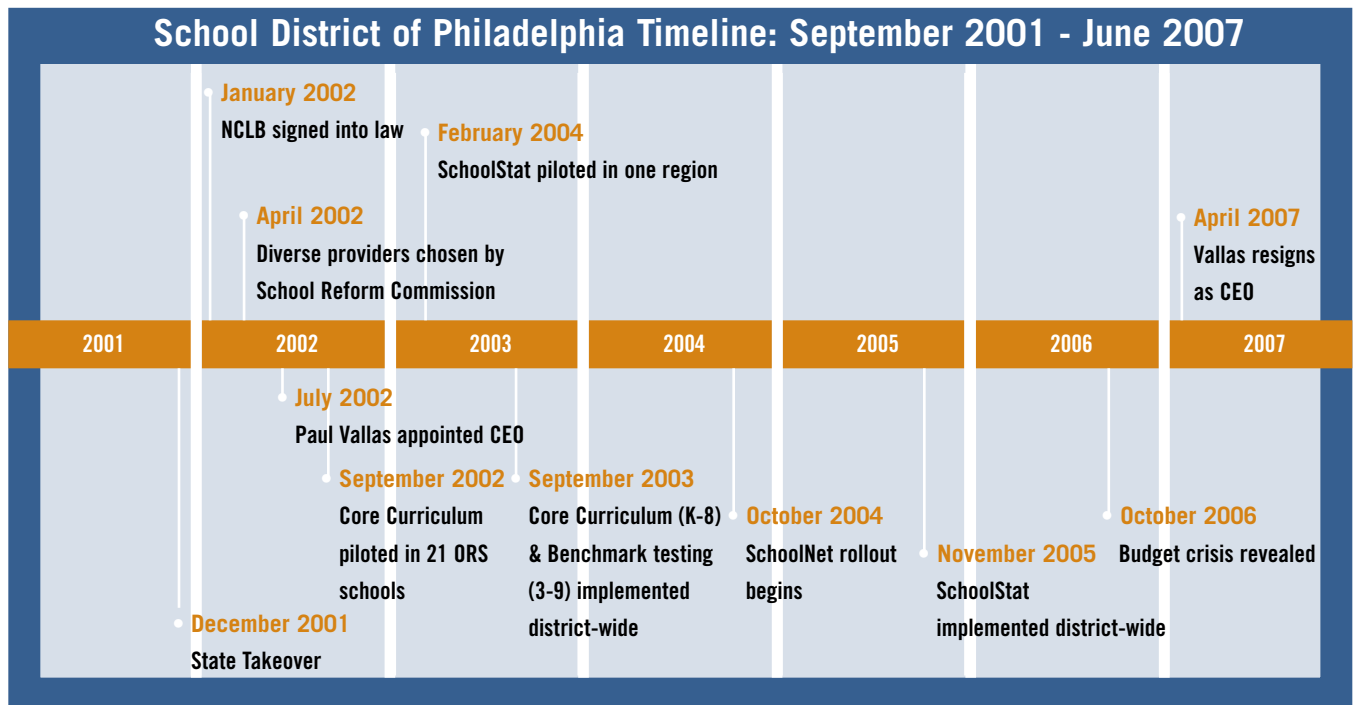


Figure 2.1

Core Curriculum

A uniform curriculum for grades K-8 in math and literacy was implemented system-wide in September 2003.

A uniform curriculum in science was implemented for grades 7 and 8 in September 2004 and implemented for grades K-6 in September 2005.

A uniform curriculum in social studies was implemented for grade 8 in September 2004 and grades K-7 in September 2005.

SchoolStat

A performance management system developed by the Fels Institute; includes

- 1) data on student performance, attendance, and school climate; and
- 2) monthly data review meetings intended to help school leaders actualize what they are learning from the data.

The SchoolStat contract was cancelled in summer 2007 in the wake of budget cuts.

Benchmarks

Interim assessments administered every six weeks to inform instruction (administered less frequently in high schools); aligned with the Core Curriculum; implemented in grades 3-9 in September 2003, and grades 10-11 in September 2004.

SchoolNet

Web-based instructional management system; includes student performance data, curricular materials, professional development materials, and online communities; users include school staff, parents, and students; about 50 schools were equipped each semester, with all schools equipped by March 2006.

Chapter Two

Philadelphia's Managed Instruction System²⁰

I tell my teachers, 'The Core Curriculum is your Bible.'
Principal

Benchmarks replace religion around here.
Teacher Leader

In response to accountability pressures from No Child Left Behind, School District of Philadelphia leaders instituted a Managed Instruction System that represented a more prescriptive approach to curriculum, instruction, and assessment than the district had taken in previous reform eras. For this chapter, we address two sets of questions: First, what were district leaders' expectations for how school staff would use Benchmark data, and what supports did they provide to help practitioners become proficient in using data to guide instruction? Second, were teachers responsive to the Managed Instruction System, particularly the Benchmark assessments? Did they use them? Did they find them helpful?

Leaders expected that data from the Benchmark assessments would be used by school practitioners in the context of a more broad-based focus on data-driven decision-making and that the data would inform planning and action at the classroom, grade, and school levels. In this chapter, we provide a description of Philadelphia's Managed Instruction System, district leaders' expectations for the use of the MIS, and the supports that were provided to help practitioners use its components. Drawing on data from the district-wide teacher survey and data from our interviews in schools, we also report teachers' responses to the MIS.

The Philadelphia Context

District-wide curriculum and student assessment has been an integral part of the School District of Philadelphia's efforts to improve education and student achievement for more than 25 years. Over this time, assessment results have been used for both instructional and accountability purposes. The centerpiece of Superintendent Constance Clayton's 12-year administration (1980-1992) was the K-12 Standardized Curriculum with a week-by-week schedule for instruction. A criterion-referenced test for each subject area administered annually measured students' mastery of the Standardized Curriculum.

²⁰ This chapter is based on a presentation by Research for Action and the Consortium for Policy Research in Education, *Building with Benchmarks: The Role of the District in Philadelphia's Benchmark Assessment System*, presented at the Annual Meeting of the American Educational Research Association, New York, NY, March 2008.

David Hornbeck, who became superintendent in 1994, brought standards-based reform to Philadelphia. The School District of Philadelphia abandoned the Standardized Curriculum of the Clayton era, shifting its emphasis from teachers covering a prescribed curriculum to all students meeting rigorous performance standards. In Philadelphia's first move towards accountability based on student achievement, the district adopted the Stanford Achievement Test (SAT9), an off-the-shelf, nationally-normed test, as an important part of the Performance Responsibility Index (PRI). Principals' performance reviews and salaries were tied to their schools' meeting district-established PRI targets.²¹ The School District of Philadelphia issued curriculum frameworks that provided teachers an overall approach to curriculum and instruction and sample lessons for different subjects and grade levels. However, the frameworks did not offer a scope and sequence, and many teachers, as well as the Philadelphia Federation of Teachers (PFT), expressed frustration with what they saw as a lack of curricular guidance.²²

Vallas had become convinced of the efficacy of a standard district-wide curriculum during his tenure as CEO of the Chicago Public Schools.

Since a state takeover of the Philadelphia school district in 2001, the district has served as a laboratory for fundamental changes in school governance and management. The most publicized of these changes was a complex privatization scheme that includes market solutions such as a "diverse provider" model of school management,²³ expansion of charter schools, and until 2007, extensive outsourcing of additional core district functions, including Benchmark assessments.²⁴ However, at the same time, the district instituted strong centralizing measures for schools that were not part of the diverse provider model.

When he came to Philadelphia in 2002, CEO Paul Vallas, with the support of the PFT, began plans for a Managed Instruction System. As shown in [Figure 2.1](#) one of Vallas' first initiatives was to institute a district-wide Core Curriculum in four academic subjects for grades K-8. Benchmark

²¹ Porter, A. C., Chester, M. D., & Schlesinger, M. D. (2002, June). Framework for an effective assessment and accountability program: The Philadelphia example. *Teachers College Record*, 106(6), 1358-1400.

²² Corcoran, T. B. & Christman, J. B. (2002, November). *The limits and contradictions of systemic reform: The Philadelphia story*. Philadelphia: Consortium for Policy Research in Education.

²³ In total, seven different organizations (three for-profit educational management organizations (EMOs), two locally based non-profits, and two universities) were hired and given additional funds to provide some level of management services in 46 of the district's 264 schools (Bulkley et al., 2004). The SRC also created a separate Office of Restructured Schools (ORS) as its own internal "provider" to oversee 21 additional low-performing schools, granted additional funding to 16 low-performing schools that were making progress (the "sweet sixteen," and converted three additional schools to charter schools (Useem, 2005).

²⁴ For example, the School District of Philadelphia contracted with Kaplan to develop the Core Curriculum for grades nine through twelve and hired outside vendors such as Princeton Review to run extensive after-school programming for students who were struggling.

assessments accompanied the Core Curriculum. Vallas had become convinced of the efficacy of a standard district-wide curriculum during his tenure as CEO of the Chicago Public Schools. Philadelphia central office staff who had served during the Hornbeck years also saw the value in this approach. They, along with staff from the Philadelphia Education Fund, developed the district's Core Curriculum for grades K-8.

Vallas made the Core Curriculum and Benchmarks mandatory for district schools that were not managed by private providers and voluntary for those managed by private providers. However, all of the providers (with the exception of Edison Schools, Inc.) adopted parts or all of the district's Core Curriculum and the Benchmark assessments.²⁵

District-Wide Teacher Survey Data Used for Analysis in this Chapter

In June 2006 and June 2007, the school district distributed a pencil-and-paper survey to all of its approximately 10,500 teachers. A total of 6,680 teachers (65 percent of all teachers) from 204 of 280 schools responded to the spring 2006 survey. A total of 6,007 teachers (60 percent of all teachers) responded to the spring 2007 survey. These response rates are comparable to that for large-scale teacher surveys in other major cities; for example, teacher surveys fielded by the Consortium on Chicago School Research typically produce a response rate of about 60 percent.

District leaders had particular expectations and theories about how teachers would use the Managed Instruction System. But how did teachers respond to it? For this chapter, we examined survey responses from elementary and middle grade teachers who said that: (a) they were teaching in a grade span in which Benchmark assessments were offered and (b) they taught either in a self-contained elementary classroom or were assigned to teach math, English, language arts, and/or reading in grade three or above. There are 1,754 teachers in the data set for 2006 and 1,941 teachers in 2007 who meet these criteria. In this report, we use the most recent data unless a particular question was not on the survey in 2007.

The Core Curriculum

In grades K-8, the Core Curriculum includes performance goals that specify what students must know and be able to do by the end of the school year, while indicating the intermediate levels of proficiency students should attain to be on track to meet state standards. The curriculum includes a specific

²⁵ Edison, Inc. was the only outside provider that came to Philadelphia with a fully-developed curriculum. It also quickly developed its own interim assessments that were designed to predict students' performance on the PSSA. When CEO Vallas heard about Edison's assessments, he decided that they were a good idea. However, curriculum and assessment staff became convinced that aligning them with the Core Curriculum was more important than having them serve a strictly predictive function.

pacing schedule that is organized by six-week instructional cycles. It indicates how many days should be spent on topics covered in the Core Curriculum and identifies the relevant textbook pages (specific textbook series are mandated for literacy, mathematics, and science). The district requires that all elementary students have 120 minutes of literacy and 90 minutes of math per day.²⁶ The Core Curriculum provides teachers with suggested “best practices” and multicultural connections that can be integrated into daily lessons. Supplemental resources for enrichment are provided, as well as strategies for working with special student populations.

It was a rare teacher who reported that he or she did not “always” or “often” use the Core Curriculum to guide instruction.

Despite these supports, the Core Curriculum poses considerable challenges for Philadelphia teachers. The district’s research-based “balanced approach” to literacy requires that teachers use guided reading groups and reading centers – instructional strategies that are new to many teachers and that test teachers’ classroom management skills. Teachers are also required to use Everyday Math (grades 1-5) and Math in Context (grades 6-8), research-based curricula developed in the 1990s and promoted by the National Science Foundation. Both math curricula emphasize problem solving and conceptual learning, an approach that challenges elementary and middle grades teachers who often do not have sufficient mathematical knowledge to choose instructional strategies that will help students scaffold from misunderstanding to understanding. These curricula also “spiral,” returning over and over again to concepts previously taught, each time developing the concept more deeply. The spiraling approach creates conflicts for teachers because, as a district administrator explained, teachers “feel uncomfortable going on [to new material] before the kids have mastered certain things.” Comments made by teachers echo this statement. For example, a third grade teacher remarked about the Everyday Math curriculum,

I just don’t believe that the children can grasp concepts in two days and then be introduced to them again three weeks later. You know, in some skills, all skills, you need consistent practice, practice with it. And I don’t believe that program gives it to them (2006).

Teachers’ Use and Perceptions of the Core Curriculum

Results from the teacher survey indicated that teachers’ responses to the Core Curriculum were generally strong and positive. By the time the district-wide teacher survey was conducted in June 2007, four years after the district-wide rollout of the Core Curriculum, it was a rare teacher (9 percent) who reported that he or she did not “always” or “often” use the Core Curriculum to guide instruction (other response choices were “occasionally”

²⁶ Travers, E. (2003, September). Philadelphia school reform: Historical roots and reflections on the 2002-2003 school year under state takeover. *Penn GSE Perspectives on Urban Education*, 2(2).

and “never”). Eighty-six percent of the teachers said that they often or always used the Core Curriculum to organize and develop course units and classroom activities. Seven out of ten teachers reported that they often or always used the Core Curriculum to “redesign assessment strategies.”

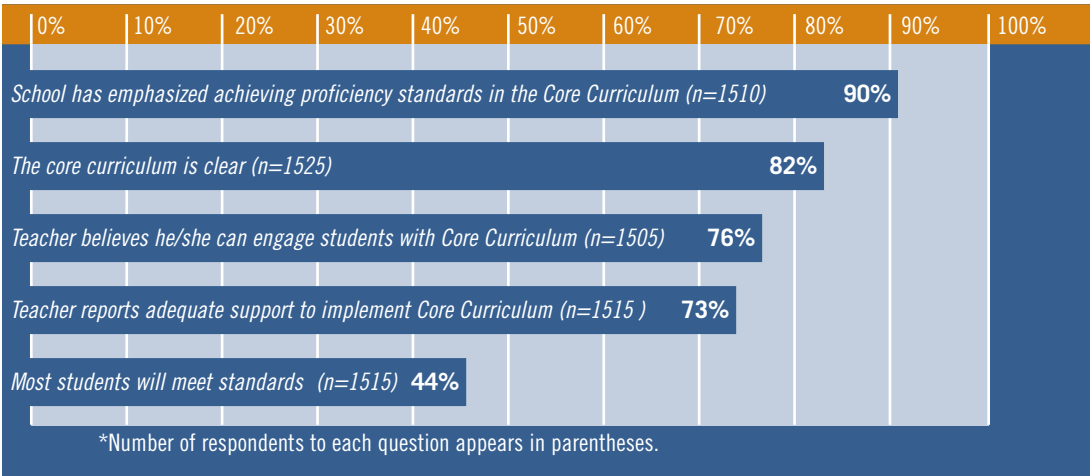
These findings are consistent with our qualitative research as many teachers were positive overall about the Core Curriculum and its ability to engage students. For example, a fifth grade teacher explained that the goal of her school was to follow the Core Curriculum with “fidelity” because it helped teachers stay on track and helped students achieve proficiency. She stated,

This year that just passed, [our goal was] to follow the Core Curriculum because we began to believe that if we followed that grade through grade that kids would be proficient. If I’m doing my own thing, you’re doing your own thing, we’re not really following one thing, the kids are not going to reach their fullest potential.
(May 2007)

Furthermore, some teachers reported making instructional changes in their classroom based on specific strategies highlighted in the Core Curriculum. They expressed confidence that using these strategies would result in increased student achievement.

As shown in **Figure 2.2**, substantial majorities of teachers reported that their school placed a strong emphasis on achieving the standards outlined in the Core Curriculum, that the Core Curriculum was clear, that they believed that they were engaging their students when implementing the Core Curriculum, and that they had received adequate support to implement the Core Curriculum. Given the teachers’ generally positive reports about the clarity of the curriculum, its capacity to engage students, and the support

Figure 2.2 Teacher Survey Responses on Core Curriculum:
Percent reporting agreement



they had received for implementation, however, it is notable that fewer than half of the teachers thought that most of their students would be able to meet the academic proficiency standards outlined in the Core Curriculum.

SchoolNet

SchoolNet is a district-wide instructional management system for the Benchmark assessments and other student data. It is intended to make assessment data immediately accessible to every classroom teacher and building principal and to provide analysis and instructional tools for educators' use.²⁷ Student information available on SchoolNet includes: PSSA and Terra Nova results (by individual, class, grade, and school), Benchmark results, student reading levels, student report card data, attendance data, and disciplinary data. (See [Table 2.1](#) for a description of the major assessments used in Philadelphia K-8 schools.) SchoolNet provides a number of other online features to assist teachers with data analysis and re-teaching, including links to the actual Benchmark items, information about how to re-teach the particular standards, and additional practice worksheets for students. To facilitate teachers' use of SchoolNet, the School District of Philadelphia planned to issue laptop computers to all teachers in district-managed schools (but not schools managed by outside providers) thus reinforcing the expectation that teachers' classroom instruction would be "data-driven."²⁸

The district expected all teachers to receive training on the use of SchoolNet and used a school-based, turnkey training approach. Generally, principals and a technology support person received professional development from the central office and were expected to return to their schools and train their staff. As one administrator described, "The principals got trained in a day during the summer. The teachers got trained on the first half day in October. The principals got the PowerPoint and the principals trained the staff. We wrote a script for them." Our research indicated that, while training did occur in the schools, there was considerable variation in whether principals' expected teachers to use SchoolNet. Several principals echoed the sentiment expressed by one, "I don't necessarily think that going on the computer to look at the data is a good use of teachers' time. We print the data for them."

Each cycle of instruction and assessment consists of six weeks: five weeks of instruction, followed by administration of Benchmark assessments and a sixth week of review and/or extended development of topics.

²⁷ Students' families also have limited access to SchoolNet data through the system's FamilyNet tool to obtain up-to-date information on their children's test scores (including Benchmark assessments), report card grades, and attendance.

²⁸ A fourth component of the Managed Instruction System was SchoolStat, a data management system that compiled and compared school level data on student performance and behavior and student and teacher attendance. Developed in partnership with the Fels Institute of Government of the University of Pennsylvania, SchoolStat was used at regular meetings of regional superintendents with their principals to discuss the status of, and ways to improve, climate and achievement in their schools. SchoolStat was discontinued in 2007, due to budget cutbacks.

Benchmark Assessments

Benchmark assessments were implemented district-wide in grades 3-8 in Philadelphia in October 2004. In the preceding two years, they had been used in the set of schools managed by the district's Office of Restructured Schools (ORS). Each cycle of instruction and assessment consists of six weeks: five weeks of instruction, followed by administration of Benchmark assessments and a sixth week of review and/or extended development of topics.²⁹

At the time of the study, the district administered Benchmarks in Reading and Mathematics to students in grades 3-8. Each Benchmark assessment was designed to test only those concepts and objectives taught since the most recent assessment was given. District leaders reported that the assessments were also aligned to Pennsylvania's assessment anchors (and, therefore, to the content of the state test) and state standards. All of the items in the Benchmark assessments are multiple choice and come directly from the concepts and skills in the district's pacing guide (called the "Planning and Scheduling Timeline"). When the Benchmarks were first implemented, students took paper and pencil tests. As schools came online with SchoolNet, students took the assessments on computers.

On the district's website, the Office of Curriculum identified multiple purposes for the Benchmark assessments (School District of Philadelphia, 2007):

- To provide PSSA practice for students by simulating rigor, types of questions and building test-taking stamina;
- To provide teachers, administrators, students, and parents with a quick snapshot of student progress;
- To determine if what is taught is what is learned;
- To help teachers reflect on instructional practices; and
- To provide data to assist in instructional decision-making.

While the district's website formally identified these purposes for the Benchmarks, analysis of interviews with central office staff suggests two central goals. First, the Benchmarks would provide feedback to teachers about their students' success in mastering concepts and skills covered in the Core Curriculum during the five-week instructional period. One district leader explained the limitations of past reliance on the state assessment PSSA for formative information,

²⁹ Journalistic accounts of the use of interim assessments (largely in *Education Week*) led us to the conclusion that in most school districts using interim assessments, the tests are given between three times a year and monthly. Aside from Philadelphia, we did not identify any other districts where time was set aside explicitly for addressing weaknesses identified from analysis of interim assessment data.

Table 2.1 District-Wide Assessments

| Assessment | Description |
|--|--|
| District Benchmark Assessments | |
| Not required in schools managed by outside providers but used in all schools in the district except schools managed by Edison Schools, Inc. | Administered at the end of the 5th week in a 6 week instructional cycle to give teachers feedback about students' mastery of topics and skills in the Core Curriculum. Reading and mathematics in grades 3-8; science in grade 3, 7 and 8. Multiple choice questions. |
| Literacy Assessments | |
| Informal reading assessments used in grades K-8. Developmental Reading Assessment (DRA) and the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) used in K-3. Gates-McGinitie used in grades 4-8 | Administered at least two times a years for the purpose of establishing students' instructional level in reading. In the early grades these assessments are administered individually and assess phonetic awareness, fluency, and re-telling. In grades 4-8 they are administered in a group setting and assess word recognition and comprehension. |
| Standardized Summative Assessments | |
| Pennsylvania System of School Assessment (PSSA) | Standards-based test in literacy, math and science used to measure achievement at district, school, grade, classroom and student level. Multiple choice and open-ended response questions aligned with Pennsylvania standards. Math and literacy in grades 3-8 and 11; science in grades 4, 8 and 11. The PSSA Writing Assessment assesses students' ability to write a five paragraph essay in response to prompt. Scored for focus, content, organization, style and conventions. Given in grades 5, 8, and 11. Not used for accountability purposes. Used in calculating whether a school makes Annual Yearly Progress under NCLB. |

We started with Benchmarks because that's the only formative piece we have. That became the one big thing that teachers had where they could change directions if they needed to make mid-course corrections. Before, you waited every year for return of the PSSA results. (2005)

Second, the six-week cycle of teaching and assessment would, as one district leader noted, “create some kind of a pacing and sequence program.”(2005) Principals and teachers confirmed that the Benchmarks provided a curriculum roadmap with specific destinations demarcated along the way. One principal described the reaction of teachers at her school: “When teachers saw kids’ results on the Benchmarks, they really knew ‘I didn’t cover this. I should have covered this.’” At another school, a fourth grade teacher remarked,

The other tests, like the tests that I give in the classroom are maybe targeting one story or one particular skill, whereas [Benchmarks] give you the big picture of what you have done in the last 6 weeks and whether you achieved what you were supposed to teach them in the last 6 weeks (2007).

Similarly, a sixth grade teacher described the Benchmarks as “checkpoints” that help him to see exactly where he is with the Core Curriculum and how well the students understand what he is teaching (2007).

Teachers’ Use and Perceptions of Benchmark Assessments

Results from the teacher survey indicated that teachers’ use of the Benchmark assessments was widespread and frequent. In 2007, fewer than three percent of teachers reported that they had never examined their students’ Benchmark assessment scores during the year. Almost half of the teachers (45 percent) said that they had examined these scores more than five times during the year, and an additional 44 percent said they had examined them three to five times. This high use held across both elementary and middle grades teachers.

The survey data indicated that a majority of teachers believe that the Benchmark assessments were a source of useful information about students’ learning. In 2006, 86 percent of the teachers reported that Benchmark assessments were useful for identifying particular curriculum topics where students still needed to improve. Likewise, in 2006, 67 percent agreed with the statement that “The Benchmark tests are a useful tool for identifying students’ misunderstandings and errors in their reasoning.” [Figure 2.3](#) presents teachers’ responses to questions about Benchmarks on the 2007 survey. Almost three quarters of the teachers said that they agreed or strongly agreed that the Benchmarks gave them a good indication of what the students were learning in their classroom (2007 data). Smaller percentages of teachers expressed posi-

tive views of the instructional consequences and pacing of Benchmarks. Sixty-one percent of the teachers felt that the Benchmark assessments had improved instruction for students with skills gaps (one of their key stated purposes), 58 percent thought that Benchmarks set an appropriate pace for teaching the curriculum, and 57 percent said that Benchmark assessments provided information about their students' learning that they would not otherwise have known – a remarkable admission for teachers to make.

These findings are consistent with our qualitative research. In our interviews with teachers, the majority reported that the Benchmarks helped them identify student weaknesses that they would have missed if they had not had Benchmark data. For example, a third grade teacher commented,

I think it really helps me to see what I need to review and go over. Okay, nobody got their fraction question right; let's go back and review fractions. It just helps me see that. (2006)

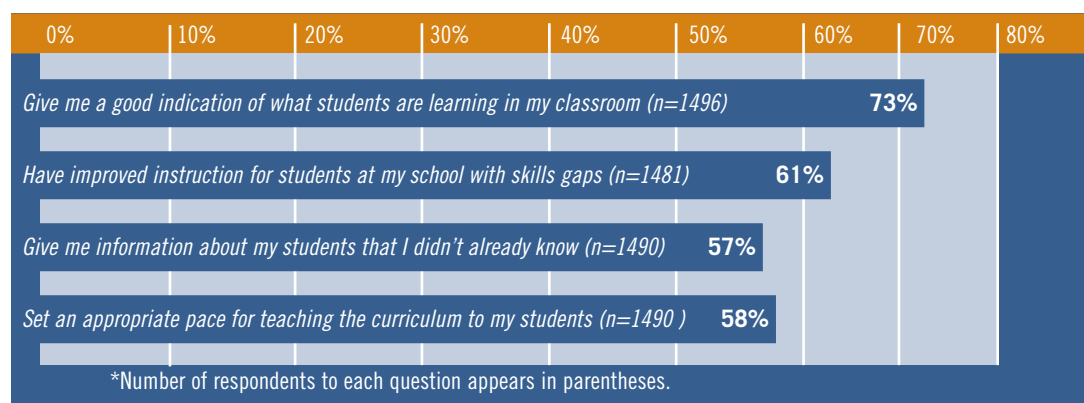
A sixth grade teacher described how she learned from the Benchmarks that her students were having difficulty following directions and needed to be shown the steps for how to complete a particular assignment.

I have to model for them how I'm thinking . . . because they weren't reading the directions and they weren't working through all the steps. (2007).

“When teachers saw kids’ results on the Benchmarks, they really knew ‘I didn’t cover this. I should have covered this.’”

- A Principal

Figure 2.3 Teacher Reports on Benchmarks:
Percentage of respondents reporting agreement



District Supports for Use of the Benchmark Data

The district provided a set of supports to all schools in the district: access to online data, resources, and reports through SchoolNet, structured tools for analyzing and reflecting on Benchmark data, and professional development. The district provided additional supports to low-performing schools.

District leaders expected individual teachers to access and use a variety of analyses of Benchmark data available on SchoolNet and to take advantage of instructional features of SchoolNet such as information about how to re-teach particular skills and concepts.

The district also developed several tools that support teachers' use of the Benchmark data: the Item Analysis Report, the Data Analysis Protocol, and the Teacher Reflection Protocol. (See **boxed text** on page 26 for a description of each of these tools.) The purpose of the Item Analysis Report is to give teachers a user-friendly way to access and manage data from Benchmark assessments. The Data Analysis Protocol, which teachers are required to hand in to principals, reinforces the expectation that Benchmarks, as a formative assessment, will be used for instructional purposes by helping teachers to think through the steps of analysis and action as they review the Item Analysis Report. District leaders expected the analysis of Benchmarks to create an opportunity for teachers to reflect on their instruction. The district leaders reasoned that, in analyzing the Benchmarks, teachers could begin to examine their own content knowledge and instructional repertoire with an eye on identifying what professional development and support would be beneficial to them. They expected teachers to use the sixth week of instruction not just to re-teach in the same old way but to find new instructional strategies that would prove more successful. One district administrator described what she hoped would be a teacher's thought process as she reviewed the Benchmark data for her class:

I think the Benchmarks give you information about your class, which then will say to you, "Okay, I've taught inference, and the Benchmarks are showing me over and over again the kids aren't getting inference. I need to do something about trying to find a resource for inference." (2005)

To encourage teachers' reflective use of the Benchmarks, the district created a single-page Teacher's Reflection Protocol intended to be completed by individual teachers following each administration of the assessment.

While the primary focus of central office staff members was on the use of Benchmark results by individual teachers, they also anticipated that various groups in the school – especially grade groups – would examine the data. The focus on groups of teachers was consistent with an emphasis on Benchmarks

District leaders expected teachers to use the sixth week of instruction not just to re-teach in the same old way but to find new instructional strategies that would prove more successful.

Tools to support teachers' use of Benchmark data

Item Analysis Report

The Item Analysis Report is generated by SchoolNet and provides teachers with an item-by-item analysis of the test at the individual student level. The Item Analysis Report provides data spreadsheets for every teacher that includes, for every student, the *correct* and *wrong* answers selected; how *many* and exactly *which* items each student answered correctly; the average percentage correct for each class for each item by state standard statement; and the state standard statement tested for each item. (A mock-up of the report can be found in Appendix B.)

Data Analysis Protocol

The Data Analysis Protocol poses the following tasks and questions:

- Using the Item Analysis Report, identify the weakest skills/concepts for your class for this Benchmark period.
- How will you group or regroup students based on the information in the necessary item analysis and optional standards mastery reports? (Think about the strongest data and how those concepts were taught.)
- What changes in teaching strategies (and resources) are indicated by your analysis of Benchmark reports?
- How will you test for mastery?

The Teacher Reflection Protocol

The Teacher Reflection Protocol includes the following writing prompts:

- In order to effectively differentiate (remediate and enrich), I need to...
- Based on patterns in my classes' results, I might need some professional development or support in...

serving instructional purposes. This expectation that teachers would talk with one another regularly was explained by a district leader who commented:

The expectation is that the 3rd grade teachers will sit at a table with each other and say, "Here's how my kids did on Item 1. How did your kids do? Whoa! My kids didn't do well. Your kids all nailed it. Tell me how you taught that? Alright, I'll go back and I'll try that." That's supposed to happen item by item. (2005)

However, it did not provide a set of tools to guide group discussions of Benchmark data. And the district professional development for principals focused on the technical aspects of accessing and organizing data, not on leading staff through conversations about the data. District leaders also expected that principals would use the Benchmark data to assess the successes and gaps in a school's instructional program. For example, the district directed principals to use Benchmark results as they developed their School Improvement Plans, a yearly exercise in which school staff assesses areas of weakness that should be a focus for improvement in the following year.

The survey results shed light on where teachers received the most help with

how to use Benchmark results. Many schools had school-based literacy teacher leaders and, less frequently, math teacher leaders. The number and mix of teacher leaders depended on availability of funding. The greatest sources of help in interpreting Benchmarks and other data and using them to make instructional decisions, according to the teachers, were the school-based literacy and math teacher leaders. One-third of the teachers reported that the literacy or math teacher leaders provided “a great deal of help,” and 76 percent said that they provided at least “some help” (possible responses were; no help, some help, and a great deal of help). Approximately two-thirds of the teachers reported that principals were at least “some help.” Clearly, school-based leaders made use of data a priority for their work with teachers. However, 69 percent of the teachers reported that regional office or central office personnel were “no help,” an indication that regional staff do not often reach classroom teachers.

Professional development for principals focused on the technical aspects of accessing and organizing data, not on leading staff through conversations about the data.

In Summary

Historically, although education reformers have had considerable success convincing districts to undertake organizational reforms, substantial instructional change in the classroom has been more difficult to achieve. This history would give good reason to suggest that teachers would look at the institution of a Core Curriculum and Benchmarks and other assessments with skepticism. However, our data from a district-wide teacher survey and qualitative research in ten schools indicated a more positive response. The Managed Instruction System is, in fact, exerting considerable influence on classroom instruction. Almost all teachers in grades 3-8 reported that they used the Core Curriculum and data from the Benchmark assessments and most found them useful. Our visits to ten schools between September 2005 and June 2007 corroborated findings from the teacher survey: use of the MIS – the Core Curriculum and Benchmarks – had permeated schools, as the quotes at the beginning of this chapter indicate.

It is likely that the historical context of the School District of Philadelphia, the district’s design of the MIS, and the supports that it implemented to help teachers use the Core Curriculum and Benchmarks contributed to teachers’ acceptance of the MIS. Philadelphia teachers were ready for the Core Curriculum and Benchmarks; they saw the value of strong curricular guidance in an era of high-stakes accountability.

The design of Philadelphia’s Benchmark assessments had two notable advantages: alignment with the Core Curriculum and the provision of another week of instruction after teachers received their students’ Benchmark results. Alignment with the Core Curriculum made Benchmark results very relevant to teachers’ instructional planning. Eighty-six percent of the teach-

ers said that they often or always used the Core Curriculum to organize and develop course units and classroom activities. Thus, alignment likely contributed to *instructional coherence* throughout the school, a key feature of schools shown to make student learning gains in Chicago and elsewhere.³⁰ *Instructional coherence* requires a common instructional framework that “guides curriculum, teaching, assessment, and learning climate” and includes expectations for student learning and teaching materials.³¹ The sixth week for remediation and extension of topics offered the opportunity for Benchmarks to serve instructional purposes by providing teachers with formative information that could guide their follow-up with students. School leaders and teachers appreciated these strengths.

Finally, the district’s infrastructure for supporting the MIS likely contributed to teachers’ acceptance of the Core Curriculum and Benchmarks. Our research showed that this infrastructure was in place by the time of this study. Most teachers reported that their school emphasized the proficiency standards in the Core Curriculum and that they received adequate support for using the Core Curriculum. Most reported that they received the Benchmark data in a timely way and that they had participated in professional development on how to access data. Additionally, from teachers’ perspective at least, school leaders had begun to organize school infrastructure to support teachers’ use of Benchmark data. Teachers reported that they had opportunities to review data with colleagues, and had received help from math and literacy teacher leaders in using data.

Our research also suggests limitations of Benchmark assessments. Districts may look to interim assessments, such as Philadelphia’s Benchmarks, for three distinct purposes – instructional, evaluative, and predictive.³² Although Perie and her colleagues note that a single assessment can serve multiple purposes, they also comment that “one of the truisms in educational measurement is that when an assessment system is designed to fulfill too many purposes – especially disparate purposes – it rarely fulfills any purpose well.”³³ Certainly, Philadelphia’s district leaders and school practitioners looked to Benchmarks for many things.

**Philadelphia’s
Benchmark assessments had two notable
advantages: alignment
with the Core
Curriculum and the
provision of another
week of instruction
after teachers
received their
students’ Benchmark
results.**

³⁰ Newmann, F. M., Smith, B., Allensworth, E., & Bryk, A. S. (2001, January). *Improving Chicago’s schools: School instructional program coherence benefits and challenges*. Chicago: Consortium on Chicago School Research.; Newmann, F. M., Smith, B., Allensworth, E., & Bryk, A. S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis*, 23, 297-321.

³¹ Newmann, F. M. et al., 2001.

³² Perie, M. et al., 2007.

³³ Perie, M. et al., 2007, p. 6

They intended for Benchmarks to serve instructional purposes by providing “results that enable educators to adapt instruction and curriculum to better meet student needs.”³⁴ As noted, the six week instructional cycle supported this intention. District leaders expected teachers to test for mastery again at the end of the re-teaching week. However, our qualitative research suggests that such teacher-developed assessment often did not occur at the end of the sixth week. It should be noted that the lack of such retesting represents a disjuncture in the steps of the feedback system described in Chapter One. Assessing the results of re-teaching is an essential part of determining whether interventions have been successful.

Other conditions, related to the assessments themselves, are also necessary in order for interim assessments to meet instructional purposes. The assessment items must not only show teachers (as well as students) what students don’t understand, but also give adequate indications of why the confusion exists, what the missteps are. The lack of open-ended questions on the Benchmark assessment was a limitation in this regard. Further, if the distracter items on a multiple-choice test are not designed well, they do not offer good clues to students’ misunderstanding. Finally, if the items operate at only the lower levels of cognition (e.g., knowledge and comprehension), and do not tap into analytical thinking, they are not good tests of conceptual proficiency.

Evaluative purposes include information about the fidelity of implementation of curriculum and instructional programs and “enforce some minimal quality through standardization of curriculum and pacing guides.”³⁵ This appears to be the greatest strength of the Philadelphia’s Benchmarks as they are currently designed.

Philadelphia’s Benchmark assessments were not designed to be *predictive of* a students’ performance on end-of-year tests. Yet, as we will show in Chapter Four, school practitioners believed that Benchmark results would predict students’ performance (and were encouraged to believe this by regional and central office staff and provider staff who worked with them). The predictive use of Benchmark results can distract school leaders and teachers from the instructional and evaluative purposes that offer the most potential for strengthening instructional capacity.

The Managed Instruction System assumed strong leadership capacity at the school level. One district leader described the principal’s complex role with regards to the professional climate that would need to be established:

The predictive use of Benchmark results can distract school leaders and teachers from the instructional and evaluative purposes that offer the most potential for strengthening instructional capacity.

³⁴ Perie, M. et al., 2007, p. 4.

³⁵ Perie, M. et al., 2007, p. 5

To give teachers the time to have the conversation to plan instruction and to support the teachers in doing what they need to do as far as giving them the resources, the professional development, the climate to feel safe to talk about what they know and what they still need to learn themselves.

School leaders needed to ensure that the school schedule accommodated grade group meetings, that these meetings were worthwhile, and that the allotted time was used to analyze and discuss student Benchmark results and to learn about new instructional techniques. It was also up to principals to help with identifying the professional development needs of their faculty, as a whole and as individual teachers, based on the results of the Benchmarks; for example, what else did teachers need to understand about the Core Curriculum? They needed to create a professional climate that encouraged professional learning through inquiry, reflection, and informed action. In Chapter Four, we delve into whether these expectations of school leaders were realistic.

In this chapter, we have established the broad acceptance of the Core Curriculum and Benchmarks by teachers and the formation of the basic infrastructure to support implementation. The next question becomes whether the Managed Instruction System, and its use of Benchmarks, had a positive impact on student achievement. We take up that question in the following chapter.

School leaders needed to ensure that the school schedule accommodated grade group meetings, that these meetings were worthwhile, and that the allotted time was used to analyze and discuss student Benchmark results and to learn about new instructional techniques.

Chapter Three

The Impact of Benchmarks on Student Achievement

An ultimate goal of systematically tracking student progress is to increase student learning. However, whether the use of Benchmark data has an actual – rather than theoretical – impact on achievement is a question that itself needs to be examined empirically. This chapter builds on analyses presented in Chapter Two, which showed that the basic infrastructure for a Managed Instruction System was firmly in place and accepted by teachers. The widespread use and acceptance of the Managed Instruction System by teachers across the school district presents an important opportunity to assess the impact of such a system on student achievement, since an essential precondition – widespread use by teachers – is met.

We asked whether students experienced greater learning gains at schools where the conditions were supportive of data use: that is, where the Managed Instruction System was more widely accepted and used and where analysis of student data was more extensive? We address this question using two types of data: student scores on standardized tests, measured over time, and data from two teacher surveys fielded by the School District of Philadelphia in the spring of 2006 and the spring of 2007.

The Organizational Learning Framework and Key Research Questions

As described in Chapter One and depicted again in [Figure 3.1](#) on page 32, the model of data use in schools posits that the organizational learning framework involves analysis of data on student learning, followed by decisions about instructional practices. When these instructional decisions are, in turn, reflected in the instruction that teachers actually deliver, increased student performance may result. In this model, then, four activities by teachers are essential to using data to increase student learning:

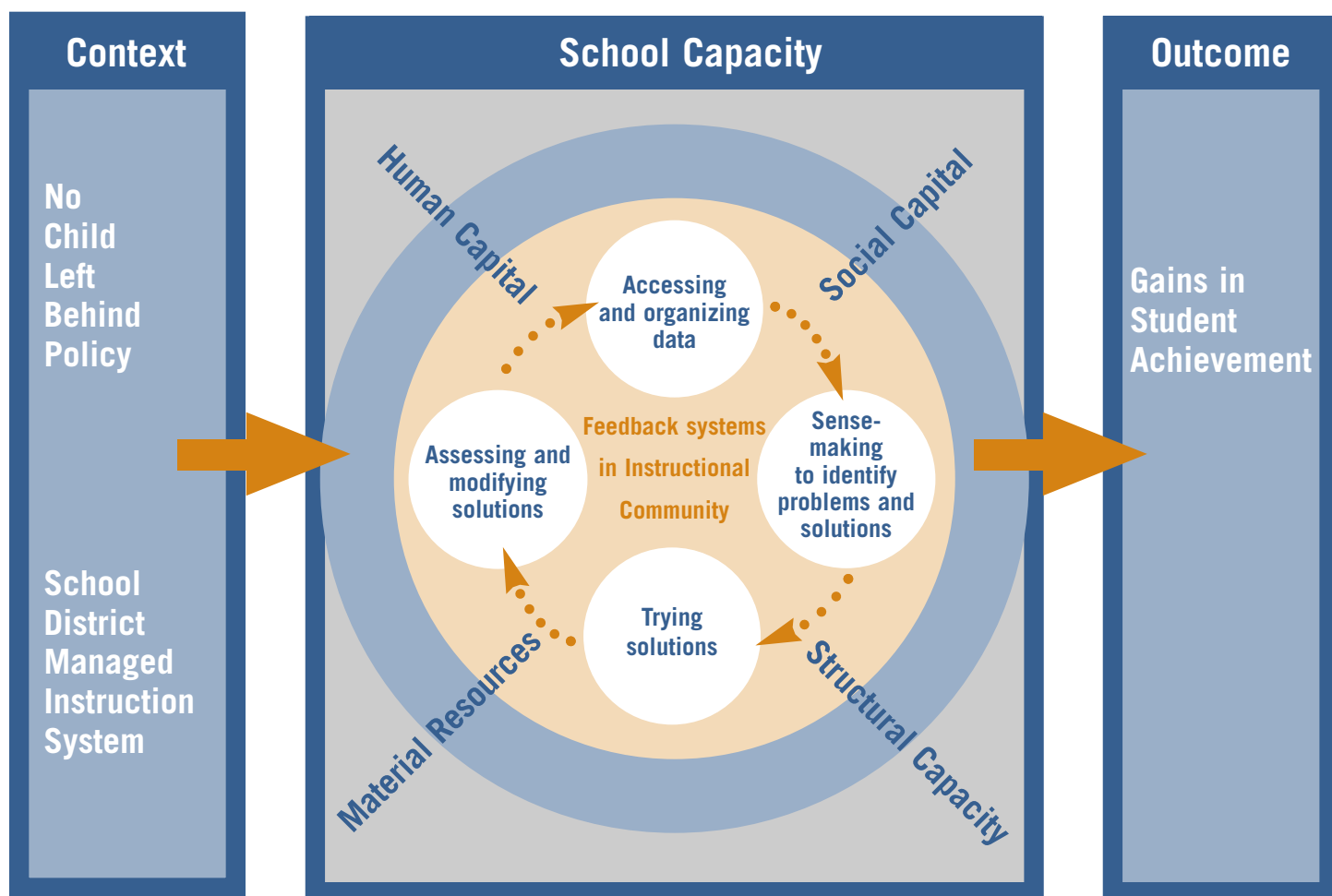
- 1) organization of data,
- 2) thoughtful analysis of student data and informed decisions about how instruction should be modified in response to the data,
- 3) faithful implementation of the instructional decisions, and
- 4) assessment of the effectiveness of instructional strategies.

The model implies that the links in the chain and the quality of the activities can affect how much students learn. The model also highlights the human, social, and material conditions – for example, the quality of leadership and relationships among staff, access to technology, professional development – that increase the likelihood of teachers being able to make good use of student data.

Documenting the skill with which teachers carry out the data analysis and subsequent instructional decisions requires a close examination of the strength of feedback systems within a school. Chapters Four and Five draw on in-depth qualitative research to explore the quality of the conversations, strategies, and decisions that arose from examining student data. Using the teacher survey data, however, we can make a broad assessment of the links between student achievement and school conditions that are fundamental for good data use in a Managed Instruction System.

Figure 3.1 depicts the organizational learning model that we incorporate into the quantitative analysis presented in this chapter. Specifically, we can examine whether teachers embraced the MIS; the availability of certain material resources for, and expertise in, examining data (human capital); the professional climate at the school (social capital and professional community); and gains in student achievement. We cannot observe the faithfulness with which teachers followed the feedback loop or the quality of their discus-

Figure 3.1 Conceptual Framework



sions, decisions, and follow-up in their classrooms. However, if we observe that student learning growth is greater at schools where conditions are more supportive of the use of a Managed Instruction System and examination of student data, then – even if we cannot examine each part of the organizational learning model – we will have preliminary quantitative evidence that examination of student data can result in greater student learning.

Analytic Approach

Our analysis relies on measurement of student academic growth, obtained from longitudinal data on student achievement made available by the School District of Philadelphia. (See the **boxed text** on page 34 for a description of how we created a measure of student academic growth.)

Data on whether conditions at schools were conducive to organizational learning that used analysis of student performance data as a driver were obtained from surveys of teachers conducted by the School District of Philadelphia during the spring of 2006 and 2007. These surveys included questions about school leadership, climate, and collegiality, developed and documented by the Consortium on Chicago School Research, as well as several sets of questions on teacher satisfaction with the Core Curriculum and Benchmark assessments, the amount of professional development for analysis of student data, access to technology that could enable viewing student data online, and collective examination of data with fellow teachers and school leaders. The scales are described briefly in the data section, below, and in more detail in Appendix E.

Our first analytic step was to examine the extent to which teachers' reports about each school condition were correlated with their reports about other school conditions. We assessed these correlations by using data at the teacher level. This descriptive work was intended to clarify whether and how school conditions tended to occur together in "packages."

Our second step was regression analysis to examine associations between student achievement and each school condition separately, controlling for individual student characteristics and the percentage of low-income students at the school. We used a two-level hierarchical linear model to analyze the relationship between student test score gains and teacher survey measures, aggregated to the school level. At Level One (the student level), we used individual-level student information to adjust for student gender, special education status, race/ethnicity, grade when taking pre-test, and grade when taking post-test. At Level Two (the school level), we controlled for the percentage of students receiving free or reduced-price lunch, using a categorical

Measure of Student Academic Growth

To create a measure of student academic growth, we examined changes in students' performance on standardized tests given at the end of successive school years. This strategy sometimes is known as a *value-added approach* because it examines the “value added” to learning by attending school in a given year. By comparing the score in the first year to the score in the second year, we obtained an estimate of how much new learning students experienced during a school year of interest. In this chapter, we examine improvement in student academic growth in two school years (2005-2006 and 2006-2007), for students in 4th through 8th grades.

To obtain a true value-added estimate, students must have taken two tests that are *vertically scaled*, meaning that the tests have been created to measure the growth in the same kinds of skills and knowledge in the same way. These vertically scaled tests become part of a family of assessments, such as the Terra Nova, Stanford Achievement Test, or, potentially, a state-developed assessment. A complicating factor for this analysis was that some of the tests students took in different years were not vertically scaled – in other words, they were part of different families of tests. To address this incompatibility between tests, we converted the student's score on each test to a ranking within the district. Students who made learning gains *relative to other students in the district* in a given year received a positive value for their learning during that year; those whose learning did not keep up with other students in the district received a negative value for the year's learning. For example, a student who scored at the 50th percentile in the district at the end of grade three and in the 52nd percentile at the end of grade four would have “moved ahead” of his peers by experiencing greater learning gains. Students who had a test score at only one point in time were excluded from the analysis.

It is essential to understand that the measure of learning that we examined is explicitly comparative. While all students could have learned something (and likely did learn) during a given school year, only students who improved their standing in the ranking of students within the School District of Philadelphia received positive scores. (For a technical description of this method, see Appendix D).

variable with four categories. More detail on the model is presented in Appendix D.

In our third step, we used multiple regression to determine the school variables that were most strongly associated with student achievement. We conducted this regression knowing from steps one and two that many of the school variables were strongly related to each other and to student achievement. What we looked for in the multiple regression were “points of leverage” – that is, school characteristics associated with higher achievement that districts could focus on in efforts to improve instruction.

Since the teacher survey was confidential, we could not link teachers’ survey responses to achievement outcomes for the specific students they taught. Therefore, for the regression analyses, we aggregated teachers’ responses to the school level, which allows us to observe the mean (average) score on particular items for each school. For example, schools with a higher mean value on an item about the quality of school leadership are interpreted as having stronger school leadership. In order to be sure that a school’s mean response was not determined by just a few staff members, we included schools in the analysis only if at least 30 percent of the teachers responded to that item. Since we could not determine the exact number of teachers in the school who taught in Benchmark subjects and Benchmark grades, we looked to see whether 30 percent of all teachers at the school responded to the survey. For this reason, we created the score for the school by using data from all teachers-respondents, rather than just those who were teaching Benchmarks.

Student Test Score Data

Student test score data from spring 2005, 2006, and 2007 were incorporated into the analysis for students who were in grades 4 through 8 during 2005-2006 and/or 2006-2007. The tests were either the Terra Nova or assessments from the PSSAs, depending on the grade and year. Raw scores for each student were converted to their percentile score within the district during the year, and these scores then were converted to standardized scores with a mean of zero and a standard deviation of one.

Teacher Survey Data

In June 2006 and June 2007, the school district distributed a pencil-and-paper survey to all of its approximately 10,500 teachers. The survey asked teachers to report on their instructional practices and use of data to inform instruction, as well as the quality of leadership, the amount of teacher collegiality, and the general climate in their school. In addition, teachers were asked about the subject(s) they taught and the grade span in which they were teaching.

A number of the survey questions were borrowed from the indicators of school leadership and climate developed by the Consortium on Chicago School Research and field-tested in surveys of teachers in the Chicago Public Schools. The indicators are described briefly below. More detail on the indicators appears in Appendix E.

Instructional Leadership

Instructional Leadership.

This indicator measures the quality of school leadership in the areas of use of student data, monitoring of instructional quality, and setting clear goals and high expectations for teachers. Since this indicator is referenced frequently throughout the rest of this chapter, it is important to note that it incorporates a number of items about the emphasis of the school leadership on using data to track student progress.

Professional Climate

Commitment to the School.

This indicator measures the extent to which teachers would prefer to work at their school than at any other school and would recommend the school to parents.

Instructional Innovation and Improvement.

This indicator summarizes teachers' reports about whether their colleagues try to improve their teaching and are willing to try new strategies.

Teacher Collective Responsibility.

This indicator measures teachers' sense of responsibility for their students' academic progress and for the overall climate of the school.

In addition, a number of survey items measured satisfaction with, and use of, elements of the Managed Instruction System. Brief descriptions follow below and detailed descriptions are provided in Appendix E.

Managed Instruction

Use of the Core Curriculum.

This measure is created from teacher reports about how much the Core Curriculum guides their topic coverage, instructional activities, and assessment strategies.

Satisfaction with Benchmarks.

This indicator measures teachers' beliefs and attitudes about whether the Benchmark assessments provide useful information about student progress in a timely and clear manner.

Collegial Instructional Responses to Student Data.

This indicator measures how often during the year teachers met with colleagues at their school to discuss re-teaching a subject or re-grouping students, based on examination of Benchmark scores.

Technology Access and Support.

This indicator measures classroom Internet access, working computers, and technology support for teachers. The indicator is not specific to the Managed Instruction System. However, student scores on Benchmarks and suggestions for instructional modifications are available on the web. Technology in good working order and support for its use would make it easier for teachers to make full use of the Managed Instruction System.

Professional Development on Data Use.

This indicator measures whether, during the school year, the school offered professional development on how to access and interpret student performance data.

Findings

Associations Among School Characteristics

Our first analytic step was to examine the correlations among three sets of variables: the measure of instructional leadership, measures of positive professional climate among teachers (teacher commitment to the school, collegial climate, and innovation), and measures of managed instruction (use of the Core Curriculum, satisfaction with the Benchmark assessments, access to technology, collegial discussions of instructional responses to student data, and professional development). These correlations, presented in [Table 3.1](#), are from the 2007 teacher survey. Only teachers who were teaching subjects and grades that used Benchmark exams are included in this correlation matrix, but the values are very similar when all teachers are included.

Table 3.1 Pearson Correlation Matrix for Key Teacher Survey Variables (2007 Survey)

| | | | | | | | | | | |
|--------------------------------------|------|------|------|------|------|------|------|------|------|--|
| Instructional leadership | 1.00 | | | | | | | | | |
| Commitment to the school | .58 | 1.00 | | | | | | | | |
| Innovation | .38 | .31 | 1.00 | | | | | | | |
| Teacher collective responsibility | .41 | .41 | .82 | 1.00 | | | | | | |
| Use of Core Curriculum | .21 | .17 | .14 | .17 | 1.00 | | | | | |
| Satisfaction with Benchmarks | .20 | .18 | .15 | .21 | .29 | 1.00 | | | | |
| Collegial instructional responses | .41 | .18 | .14 | .18 | .23 | .33 | 1.00 | | | |
| Technology access and support | .31 | .32 | .25 | .26 | .12 | .15 | .14 | 1.00 | | |
| Professional development on data use | .28 | .18 | .10 | .10 | .16 | .09 | .23 | .10 | 1.00 | |

Instructional Leadership

Professional Climate

Managed Instruction

Instructional leadership

Commitment to the school

Innovation

Teacher collective responsibility

Use of Core Curriculum

Satisfaction with Benchmarks

Collegial instructional responses

Technology access and support

Professional development on data use

There are moderate-to-strong positive associations within the group of variables that speak to *instructional leadership* and *positive professional climate among teachers* (teacher commitment to the school, collegial climate, and innovation). For example, the correlations between instructional leadership, on the one hand, and the professional climate variables, on the other, range from .38 to .58. Further, the correlations among the three variables that address professional climate are particularly strong, ranging from .41 to .82. Finally, and importantly, the correlation matrix also shows that strong instructional leadership and a positive professional climate are positively associated with the five “managed instruction” variables.

A reasonable conclusion from these correlations is that the school characteristics of strong instructional leadership, a positive professional climate, investment in the Managed Instructional System, and use of student data to inform instruction tend to be found together. That is, they co-occur as “packages” because schools that are “good” in one respect tend to be “good” in other respects; schools with strong instructional leadership are often schools where teachers trust each other and encourage their colleagues to innovate and grow professionally. From a research perspective, these characteristics of schools can be difficult to separate analytically, requiring us to choose one variable to serve as a proxy for a range of favorable conditions at the school.

That said, it is notable that of the four variables that describe school leadership and professional climate, *instructional leadership* has the strongest relationship with the five variables related to the Managed Instruction System. For example, the correlation for instructional leadership and the frequency with which teachers met to discuss instructional responses to student data is .41, while the correlation between innovation and discussion of instructional responses to data is just .14. It is worth recalling that, in this study, instructional leadership refers to the extent to which the school leadership emphasizes data-driven decision-making, tracks student progress, knows what kind of instruction is occurring in classrooms, and encourages teachers to use what they learn from professional development. It makes sense, then, that *instructional leadership*, defined in this way, would be a good predictor of how often teachers met to discuss instructional responses to student data (the *collective examination* variable) as well as the amount of professional development provided on topics related to student data.

Our model of organizational learning posits that the quality of school leadership is an important factor that supports “take-up” of the Managed Instruction System and collective examination of student data. It is not difficult to imagine that *instructional leadership* would be an important condition that would allow innovation and collegial learning – including analysis

The school characteristics of strong instructional leadership, a positive professional climate, investment in the Managed Instructional System, and use of student data to inform instruction tend to be found together. They co-occur as “packages” because schools that are “good” in one respect tend to be “good” in other respects.

of student data – to operate. The moderate or strong relationship between instructional leadership and every other variable presented in [Table 3.1](#) supports this argument. Further, the centrality of the *instructional leadership* variable to effective data use by faculty is shown in subsequent analyses in this chapter.

Also of note is that among the five MIS variables, the highest correlations are between perceptions of the usefulness of Benchmark assessments and frequency of examination of student data with colleagues ($r=.33$) and usefulness of Benchmarks and use of the Core Curriculum ($r=.29$). The first correlation supports the idea that learning from data is a social activity.

Benchmark data are useful to teachers when they have opportunities to discuss them with colleagues. The second correlation indicates the mutually reinforcing relationship between the Core Curriculum and the Benchmarks that the district intended. The more teachers invest in the Core Curriculum by adhering to it, the more useful Benchmark assessments are likely to seem as a tool to guide instruction, since the Benchmarks are aligned with the Core Curriculum. The reverse is also likely to be true: the more a teacher finds results from Benchmark assessments to be informative, the more willing he or she is likely to adhere to the Core Curriculum.

Relationships between School Characteristics and Achievement

The preceding section emphasized the positive relationships among instructional leadership, a positive professional climate, use of key elements of the Managed Instruction System, and support for teachers' use of the student data. In this section, we use a multilevel model to examine the relationships between each of these variables (aggregated to the school level) and growth in student learning. Since the *instructional leadership*, *professional climate*, and MIS variables are so inter-related, we examine separately the association between each variable and student achievement growth. Beginning on page 42, we identify and discuss the school variables that are the strongest and most consistent predictors.

[Table 3.2](#) presents the coefficients from separate multilevel regressions predicting mathematics and reading growth in 2005-2006 and 2006-2007. Thirty-six separate regressions are represented in the table. The variables are standardized so that the magnitude of the effects can be compared.

There are several important patterns to note in [Table 3.2](#). First, almost every variable is a statistically significant predictor of learning growth. Second, there is a positive relationship between all of the school variables and student learning growth. Schools where teachers reported stronger

Learning from data is a social activity.

Benchmark data are useful to teachers

when they have opportunities to discuss

them with colleagues.

Table 3.2 Relationships between Student Learning Growth and School Variables

| | Reading 2005-06 | | Math 2005-06 | | Reading 2006-07 | | Math 2006-07 | |
|--|-----------------|-------|--------------|-------|-----------------|-------|--------------|-------|
| | Estimate | p* | Estimate | p | Estimate | p | Estimate | p |
| Instructional Leadership | 0.11** | 0.000 | 0.12 | 0.000 | 0.17 | 0.000 | 0.15 | 0.000 |
| Commitment to the School | 0.18 | 0.000 | 0.18 | 0.000 | 0.17 | 0.000 | 0.14 | 0.000 |
| Instructional Innovation & Improvement | 0.20 | 0.000 | 0.20 | 0.000 | 0.15 | 0.000 | 0.16 | 0.000 |
| Collective Responsibility | 0.19 | 0.000 | 0.18 | 0.000 | 0.14 | 0.000 | 0.15 | 0.000 |
| Use of the Core Curriculum | 0.18 | 0.000 | 0.14 | 0.001 | 0.13 | 0.002 | 0.09 | 0.040 |
| Collegial Instructional Responses | 0.13 | 0.000 | 0.11 | 0.001 | 0.03 | 0.510 | 0.03 | 0.530 |
| Technology Access and Support | 0.15 | 0.000 | 0.14 | 0.000 | 0.10 | 0.000 | 0.08 | 0.001 |
| Professional Development on Data Use | 0.13 | 0.010 | 0.14 | 0.007 | 0.14 | 0.001 | 0.13 | 0.006 |
| Satisfaction with Benchmarks | 0.04 | 0.380 | 0.02 | 0.650 | 0.07 | 0.078 | 0.07 | 0.140 |

*The p-value is the probability that the estimate is simply the result of chance.

** Statistical significance is indicated in bold type.

instructional leadership, a more positive professional climate, greater use of the Core Curriculum, and more supports for data use by teachers experienced greater learning gains than schools without the same positive features. The effects of the school variables are observed even after controlling for individual student characteristics (demographics, special education or English Language Learner status, and grade in school) and the percentage of students at the school who were from low-income families.

In [Table 3.2](#), the coefficients range approximately from .10 to .20 for each year and each subject. Generally speaking, the instructional leadership and professional climate variables have slightly larger impacts on achievement than the MIS variables, although the magnitudes of the effects are quite close. For example, for reading growth during the 2006-2007 school year, the magnitude of the effect for instructional leadership was .17, in contrast to .10 for technological access and support and .13 for use of the Core Curriculum. An effect of .17 is considered to be of moderate size in education research.³⁶ That is, for each one standard deviation increase in the mean reported quality of the school's instructional leadership, the school's achievement ranking in the district was predicted to increase by .17 of a standard deviation.

³⁶ Lipsey, M. W., and Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48, 1181-1209.

There are two variables that, at least in some years, do not have statistically significant associations with achievement growth. A measure of *satisfaction with Benchmarks* was not significantly associated with either reading or math achievement growth, for either 2005-2006 or 2006-2007 (although it approached statistical significance at $\alpha=.05$ in 2006-2007). Likewise, a measure of *collegial instructional responses to student data* was not a significant predictor in 2006-2007. The direction of the coefficients was positive in all cases.

A measure of satisfaction with Benchmarks was not significantly associated with either reading or math achievement growth.

The framework that informs this study may provide some insight on the weak relationship between *satisfaction with Benchmarks* and achievement. The framework hypothesizes that the link between the data itself and student achievement is moderated by interpretation, subsequent instructional decisions, implementation of those decisions, and assessment of those decisions. The measure of *satisfaction with Benchmarks* tells us about only a small piece of that process: whether the teachers felt that Benchmarks provided useful, clear, and timely information about student progress. It does not tell us whether teachers had good ideas about how to respond to the data. Although accessing clear data in a timely way is important, it is insufficient for producing student achievement. As the case studies of the next chapter show, the ability of teachers to make sense of the data and plan appropriate instructional responses is heavily contingent on school resources, especially the quality of leadership and support provided by the principal and content area teacher leaders. It is also possible that there were inadequacies in the quality of the Benchmark assessments that lead to a weak relationship between teachers' satisfaction with the Benchmarks and gains in student achievements. As stated in the Introduction, a review of the technical quality of the assessments was beyond the scope of this study.

Identifying the Strongest Predictors of Achievement

In our final step, we used multivariate regression to identify school characteristics that had an especially strong relationship with achievement. Our purpose in so doing was to assess whether there were particular organizational characteristics on which education leaders could focus in order to help teachers make the most of student data.

When the relative strength of the four *instructional leadership* and *school climate* variables was tested in multiple regressions, the two variables that had the strongest and most consistent relationships with student achievement across years and subjects were *instructional leadership* and *teacher collective responsibility*. We then added each of the five MIS variables to a regression with either the instructional leadership or collective responsibility

measures. One of these MIS variables – *use of the Core Curriculum* – was a statistically significant predictor of student achievement growth in some years and for some subjects.

Table 3.3 presents the results of two regressions that include use of the Core Curriculum along with instructional leadership and collective responsibility, respectively. When *instructional leadership* and *use of the Core Curriculum* are included together as predictors of achievement, the magnitude of the *leadership* effect ranges from .08 to .15; the *Core Curriculum* effect is significant for reading and mathematics in the 2005-2006 school year; and the r-squared ranges from .06 to .12. The magnitudes of the effects and the r-squared are similar for a regression that includes *collective responsibility* and *use of the Core Curriculum*. Substantively, these regressions suggest that schools with stronger instructional leadership, a stronger sense of collective responsibility among teachers, and/or greater use of the Core Curriculum to inform content, instruction, and assessment produced greater student learning gains than other schools.

None of the other Managed Instruction System (MIS) variables was a significant predictor of achievement growth when entered into a regression with instructional leadership or teacher collective responsibility.

Schools with stronger instructional leadership, a stronger sense of collective responsibility among teachers, and/or greater use of the Core Curriculum to inform content, instruction, and assessment produced greater student learning gains than other schools.

Table 3.3 Key School Variables Predicting Growth in Student Learning

| | Reading 2005-06 | | Math 2005-06 | | Reading 2006-07 | | Math 2006-07 | |
|--|-----------------|-------|--------------|-------|-----------------|-------|--------------|-------|
| | estimate | p | estimate | p | estimate | p | estimate | p |
| Instructional Leadership | 0.08* | 0.010 | 0.10 | 0.002 | 0.15 | 0.000 | 0.15 | 0.000 |
| Use of the Core Curriculum | 0.15 | 0.002 | 0.10 | 0.030 | 0.04 | 0.300 | .00 | 0.976 |
| <i>R-squared at Level 2 (school level)</i> | .08 | | .06 | | .12 | | .09 | |
| Collective Responsibility | 0.17 | 0.000 | 0.17 | 0.000 | 0.13 | 0.000 | .14 | 0.000 |
| Use of the Core Curriculum | 0.12 | 0.004 | 0.08 | 0.060 | 0.08 | 0.053 | .03 | 0.476 |
| <i>R-squared at Level 2 (school level)</i> | 0.13 | | .10 | | .09 | | .07 | |

* Statistical significance is indicated in bold type.

In Summary

In this chapter, we discussed the results of our efforts to disentangle the impact of various factors on growth in student achievement. Importantly, we found that some factors were stronger and more consistent predictors of achievement gains than others. In particular, we found that instructional leadership and collective responsibility were strong predictors of learning growth. Use of the Core Curriculum was also a robust predictor, showing more power in 2005-06 and in reading than in math. The implications of these findings, we suggest, are powerful. In particular, we suggest that translating student data into student achievement requires a strong learning community at the school. The instructional leadership and collective responsibility measures imply that school leaders and faculty feel accountable to one another, that they are diligent in monitoring student progress, and that they are willing to use data as a starting point for inquiry.

It is notable that these measures of school leadership and school community are stronger predictors of student learning growth than satisfaction with the usefulness of Benchmark data. While Benchmarks may be helpful, they are not in themselves sufficient to bring about increases in achievement without a community of school leaders and faculty who are willing and able to be both teachers and learners.

While Benchmarks may be helpful, they are not in themselves sufficient to bring about increases in achievement without a community of school leaders and faculty who are willing and able to be both teachers and learners.

Chapter Four

Making Sense of Benchmark Data

The quantitative analysis presented in Chapter Three established that strong instructional leadership and collective responsibility were the most robust predictors of growth in student achievement, with use of the Core Curriculum being slightly less robust. It also highlighted the difficulty of analytically separating individual characteristics of schools such as instructional leadership, professional climate, use of the Core Curriculum, and use of student data to inform instruction. These characteristics tended to co-occur as “packages.”

In this chapter we use our *qualitative* data to uncover what school leaders – principals and teacher leaders – actually do as they work with teachers in instructional communities to make sense of Benchmark results and plan instructional actions. We wanted to determine, what can school leaders do to ensure that the use of Benchmark data contributes to organizational learning and ongoing instructional improvement within and across instructional communities?

In theory, instructional communities, such as grade groups, provide “an ideal organizational structure” for school staff to learn from data and use data to improve student learning.³⁷ “Organized talk”³⁸ in instructional communities is foundational for building shared understanding of issues and concerted efforts to remedy problems. In the four-step feedback system described in Chapter One, organized talk is represented in the second step, “sense-making with data to identify problems and solutions.” (See [Figure 4.1](#)) School leaders have a key role to play in facilitating interpretation of data to create actionable knowledge.³⁹ But few studies of schools have looked closely enough at how school leaders facilitate collective interpretation of data in instructional communities – what do practitioners talk about and how do they talk about it. We use our observations of grade group meetings to examine and assess the quality of interpretation processes and the factors that influenced that quality.

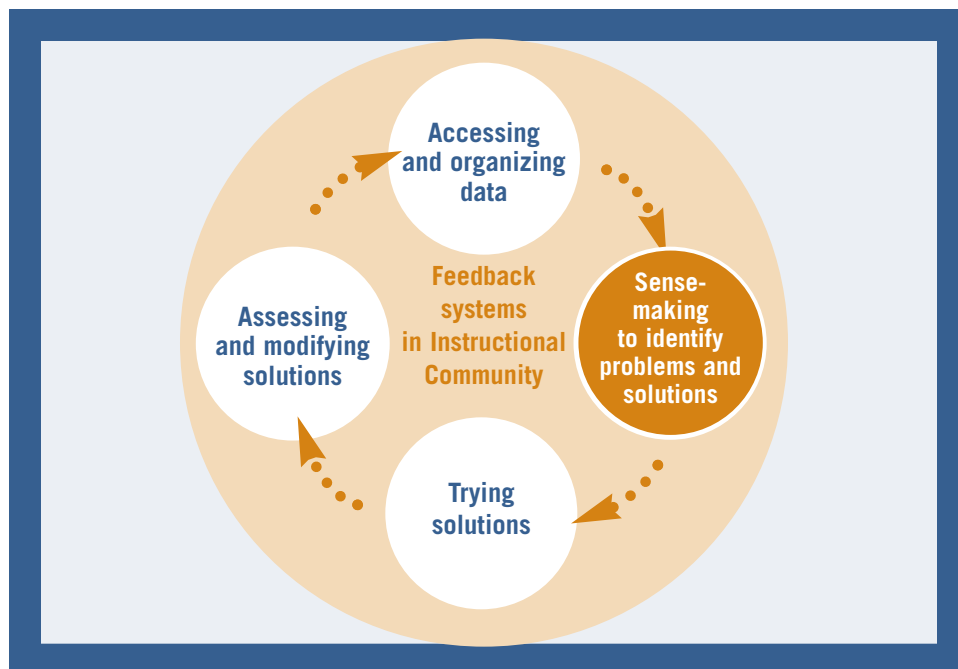
Few studies of schools have looked closely enough at how school leaders facilitate collective interpretation of data in instructional communities – what do practitioners talk about and how do they talk

³⁷ Mason, S. A. & Watson, J. G. 2003.

³⁸ Rusch, E. A. (2005). Institutional barriers to organizational learning in school systems: The power of silence. *Educational Administration Quarterly*, 41, 83 – 120. Retrieved on May 8, 2007, from SAGE Full-Text Collections.

³⁹ Daft, R. L. & Weick, K. E. (1984). Towards a model of organizations as interpretation systems. *Academy of Management Review*, 9(2), 284-295.

Figure 4.1 Feedback Loop for Engaging with Data



Strategic sense-making focuses on the identification of short-term tactics that help a school reach its Adequate Yearly Progress (AYP) targets.

Three Kinds of Sense-Making: Strategic, Affective, and Reflective

Our observations of grade groups suggest that practitioners engaged in three major types of sense-making as they sat together to discuss and interpret Benchmark data: strategic, affective, and reflective. Not surprisingly, the pressures of the accountability environment strongly influenced their sense-making. However, our observations also showed that the actions of school leaders could mediate these policy forces to create instances of substantive professional learning for school staff. Disappointingly, such instances were infrequent. There is an important opportunity for the district to strengthen the impact of Benchmark data on teacher and student learning. Below, we discuss the three kinds of sense-making.

Strategic sense-making focused on the identification of short-term tactics that help a school reach its Adequate Yearly Progress (AYP) targets. Strategic sense-making included conversations about “bubble students” who have the highest likelihood of moving to the next level of performance (from Below Basic to Basic or from Basic to Proficient) thereby increasing the probability that the school would meet its AYP goal. These conversations related to the predictive purpose of interim assessments in the framework offered by Perie et al.,⁴⁰ described in the Introduction. Strategic conversations also focused on improving test-taking conditions and test-taking strategies.

⁴⁰ Perie, M. et al., 2007.

Three Kinds of Sense-Making: Strategic, Affective, and Reflective

Strategic Sense-Making: *Most Common*

Focuses on short-term tactics that help a school reach its Adequate Yearly Progress targets, including having conversations about students who have the highest likelihood of moving to the next performance level.

Affective Sense-Making: *Common*

Focuses on teachers' professional agency and responsibility, beliefs about their students, desire to encourage one another, and motivate their students.

Reflective Sense-Making: *Least Common*

Focuses on questioning and evaluating the instructional practices used in the school and what teachers need to learn in order to help students succeed.

Finally, in strategic conversations, practitioners used Benchmarks for evaluative purposes as they worked to identify strengths and weaknesses that cut across grades and classrooms so that they could allocate resources (staff, materials, and time) in ways that increased the odds that the school would meet its AYP goal (e.g., assigning “strong” teachers to the accountability grades, purchasing calculators, lengthening instructional time for literacy and mathematics). In our observations, strategic sense-making dominated the talk about Benchmark data.

Affective sense-making included instances in which leaders and classroom teachers addressed their professional agency, their beliefs about their students, their moral purpose, and their collective responsibility for students’ learning. During affective talk, school leaders and teachers offered one another encouragement. They expressed a “can do” attitude, often relating this sense of professional agency back to the pressures that they felt from the accountability environment. In affective talk, practitioners also affirmed their belief that their students “can do it.” They discussed how to motivate their students to put forth their best effort on standardized exams and in general. Affective sense-making was the second most prevalent kind of discourse that we observed.

Reflective sense-making occurred when teachers and leaders questioned and evaluated the instructional practices that they employed in their classrooms and their school. They connected what they were learning about what their

students knew and did not know to key concepts in the Core Curriculum and they identified resources that would help them strengthen instruction of those concepts. Researchers have pointed out the importance of reflective discourse as “a springboard for focused conversations about academic content that the faculty believes is important for students to know.”⁴¹ These conversations helped teachers focus on what they needed to learn in order to help their students succeed. Such discourse about the curriculum served to shift teachers’ attention away from students’ failures and towards analyzing and strategizing about their own practices.

Reflective sense-making offers the most promise for building instructional capacity because it focuses on teachers’ learning.

In summary, reflective conversations helped practitioners plan the kinds of professional development that would strengthen teachers’ understanding and use of the Core Curriculum. They generated consideration of what other kinds of data they needed to take into account as they made sense of the Benchmark results. They offered the most promise for building increased school and classroom instructional capacity.

Making Sense of Benchmark Data: Four Examples

Below, we use fieldnotes from observations of grade group meetings in four schools to construct descriptions of the typical processes of school leaders and grade groups as they made sense of Benchmark data. These grade group meetings were consistent with what teachers and school leaders told us about their use of Benchmark data in interviews and with other types of meetings that we observed. The examples provide windows into why instances of strategic and affective talk were so prevalent. They also shed light on why the survey variable, *teacher satisfaction with Benchmarks*, was not associated with gains in student achievement. Finally, they suggest opportunities for increasing instances of reflective conversations about Benchmark results as a springboard for staff to learn more about their students, the curriculum, and pedagogy.

Attendance at each of the four meetings that we describe below consisted of the school’s principal, at least one teacher leader (usually a reading or math coach), and between two to four classroom teachers.⁴² In the four schools,

⁴¹ Mintz, E., Fiarman, S. E., & Buffett, T. (2006). Digging into data. In K. P. Boudett, E. A. City, & R. J. Murname (Eds.), *Data wise: A step-by-step guide to using assessment results to improve teaching and learning* (81-96). Cambridge, MA: Harvard Education Press, p. 94.

⁴² In order to minimize some aspects of variation and to focus on different types of sense-making relative to Benchmark data, these examples are drawn from a small subset of observations conducted between January 2005 and December 2006 in which the organizational context of the observations (grade group meetings) and the tools (the Benchmark Item Analysis Report) were held constant.

grade group meetings generally occurred every week or every other week and involved teachers from the same grade or from consecutive grades (K-2, 3-5). In each of the examples, school leaders and teachers were using the district's Item Analysis Report available on SchoolNet. (See page 26 for a description of the Item Analysis Report.) In some grade groups, principals played particularly prominent roles, but in every grade group, teacher leaders, and to a lesser extent, classroom teachers, also were active participants.

Sense Making Example 1: Encouraging re-teaching to emphasize procedures for multi-step math problems

The principal opened the discussion of the Benchmark data by asking: "How many students are Proficient or Advanced? How many are close to Proficient or Advanced? What are the questions that gave the students the most problems?" Teachers took time to use colored highlighters to note students' different status and to make decisions about tutoring assignments.

A 4th grade teacher pointed out that most of her students missed a question about the length of a paper clip because they didn't notice that the paper clip was placed at the 2 cm mark on the ruler in the picture, not at 0: "They needed to subtract 2 to get the right answer." The math teacher leader reassured the 4th grade teacher that "It's the evil test makers at work. Nobody ever starts measuring something from 2 cm."

The principal chimed in with sympathetic comments about test questions that defy common sense. She also reminded the teachers that re-teaching can be an opportunity to point out what students must keep in mind as they approach test items on the Benchmark and PSSA tests. "The re-teaching opportunity can be powerful, especially if it's done right after students take the test and it is fresh in their minds. Sometimes it's two or three steps (in a math problem) that you need to get to in order to get the right answer."

Later in the meeting, the principal offered to teach a lesson about fractions and decimals to the 4th graders, another concept that had stumped many students.

Many of the meetings we observed began in the same way that this one did, with the principal or a teacher leader asking: "How many students are Proficient or Advanced? How many are close to Proficient or Advanced?"

Even though the Benchmark data are meant to provide *diagnostic* information about what students have learned in the previous five weeks, conversations about results often assumed that they were *predictive* of performance on the PSSA – evidence of how the state's accountability measure pervaded practitioners' thinking about what they could learn from the Benchmark

"The re-teaching opportunity can be powerful, especially if it's done right after students take the test and it is fresh in their minds."

- A Principal

data. Practitioners from all of the schools in our qualitative sample reported that the identification of bubble students – students on the cusp of scoring Proficient or moving from Below Basic to Basic – was a common practice in their analysis of Benchmark data.

The teachers put stars next to those kids that they're going to target. And we made sure that those kids had interventions, from Saturday school to extended day, to Read 180. And then we followed their Benchmark data. Those were the kids that the teachers were really going to focus on, making sure that those kids become Proficient, or move that 10 percent out of the lower level so that we can make Safe Harbor next year. (Teacher, 2006)

School leaders reported that they were encouraged by the district and provider staff who worked with their schools to pay attention to proficiency levels and to track the progress of students who would be most likely to score proficient with additional supports.

The principal in this example implored teachers to strike while the iron was hot and take advantage of the re-teaching opportunity immediately so that students could see where they went awry – a strategy that research on formative assessment recommends.⁴³ And, in fact, all of the teachers at this school made a practice of going over responses to assessment items with their class right after they finished the test. In this example, however, the principal focused on re-teaching the *procedural* aspects of the math problem (“sometimes it’s two or three steps that you need”), rather than returning to the *concepts* under study – a point that we will take up again in Example 2.

Sense Making Example 2: Identifying motivational strategies and tutoring resources

At this school, the 5th grade teachers said that their students were having a lot of difficulty with Benchmark items related to fractions, particularly reducing improper fractions. One teacher noted that she had connected fractions to a lesson that she had done earlier and that, “A lot of light bulbs went off [when students saw how to draw on what they already knew].” Building on this, the principal said that she loved the image of students “tapping into prior knowledge” and suggested that everyone make posters of light bulbs for their classroom to motivate students during the Benchmarks and other tests. “Tell students to hang up a light bulb, put on your thinking caps and say ‘I can do it.’” The principal also pointed out that their volunteer tutors might be a good resource to help students who were having trouble with fractions.

The principal encouraged teachers to help their students believe they “can do it” – an example of affective sense-making in which school-based practitioners focus on how to motivate their students.

⁴³ Black, P. & Wiliam, D. 1998.

In this example, the principal diverted the conversation to address how to motivate students. She encouraged teachers to help their students believe they “can do it” – an example of affective sense-making in which school-based practitioners focus on how to motivate their students.

As in the previous example, no one in the meeting addressed conceptual issues related to mathematical content. Students were challenged by items related to fractions, but the conversation did not explore the intended purpose of these questions. As Spillane and Zeuli (1999)⁴⁴ found in their study of mathematics reform, our research indicates that discussions about Benchmark data most often did not focus on building teachers’ “pedagogical content knowledge.”⁴⁵

Pedagogical content knowledge couples knowledge about content to knowledge about pedagogy. Teachers with strong pedagogical content knowledge understand what teaching approaches fit the content being taught; their deep understanding of content makes it possible for them to explain disciplinary concepts to students and to craft learning tasks that build students’ conceptual understanding; their broad repertoire of instructional strategies provide them with options to help students with different learning needs. The alignment of Benchmark assessments with the Core Curriculum offers the opportunity for teachers to look at results with an eye towards strengthening their pedagogical content knowledge. Our observations of grade group meetings and our interviews with school leaders indicate that this was rarely a focus of practitioners’ analysis.

Sense Making Example 3: Revamping classroom routines to support student independence

The math teacher leader suggested that middle grade students need more independence during regular classes in order to improve their performance on tests. “One of the reasons that people say the kids know the material, but don’t test well, is that the conditions are so different. During instructional periods, you need to let the kids do more on their own, so it’s more like a testing situation where they have to interpret the instructions on their own.”

He suggested that the teachers should tell students the objective for the lesson, then have them work in small groups to figure out what is being asked of them in the directions for the math activity. Teachers

Discussions about Benchmark data most often did not focus on building teachers’ “pedagogical content knowledge.” Deep understanding of content makes it possible for teachers to explain disciplinary concepts to students and to craft learning tasks that build students’ conceptual understanding.

⁴⁴ Spillane, J. P. & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. *Educational Evaluation and Policy Analysis*, 21(1), 1-27.

⁴⁵ Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1-22.

should circulate during this time, noting where students are on the right track and where they are not. They should ask questions that will help students improve their interpretations. He concluded, “Our students need to learn to be more independent. After they’ve finished the task, then you can review and reflect with the small groups about how it went.”

Like the principal in the first example, this math teacher leader offered to come into classes and help teachers if they were ready to try out some of the new instructional practices discussed.

The math leader in this example made the broad point that students need to learn to work more independently and then offered specific ideas for doing this. Although these suggestions were meant to address problems students encounter in the testing situation, they are also good instructional practice.

Offers of support from school leaders are prominent in Examples 1, 2, and 3, as are teaching tips. Principals and teacher leaders offered to conduct demonstration lessons and to consult about classroom management of small groups. They also suggested steps that teachers might themselves take – re-teaching, a change in classroom routines that would encourage more student independence, ways to motivate students. We read many of these offers of support and recommendations as ways for school leaders to demonstrate their investment in teachers’ struggles and to encourage teachers in the context of the larger accountability policy context that often stigmatizes schools, educators, and students for low student achievement rather than supporting and rewarding them.

Our interviews of staff suggest that follow-up by principals and teacher leaders in classrooms was much less likely to occur in most schools than one might hope, a gap that weakens the kinds of feedback systems necessary for organizational learning. When leaders do not visit classrooms to see whether teachers are trying the strategies discussed in grade group meetings and whether they use the strategies well, an important evaluative function of Benchmark assessments is lost. Leaders do not have good information to judge the efficacy of the solutions.

Sense Making Example 4: Understanding the standards and learning how to teach standards-based content

At a fourth school, teachers brought the Item Analysis Report for their classrooms as well as copies of the Core Curriculum, having already made notes to themselves about student strengths and weaknesses. When teachers brought up the difficulty their students were having with reading the math problems on the Benchmark assessment, the principal reminded them that they could read the math questions to students.

Follow-up by principals and teacher leaders in classrooms was much less likely to occur in most schools than one might hope, a gap that weakens the kinds of feedback systems necessary for organizational learning.

The principal directed these fourth grade teachers to think about the relationship between the Benchmark assessments and the Core Curriculum standards in order to figure out why some questions were presenting more difficulty for students than others. “Look at questions that test the same standard. Are they written the same way or a different way? Is one harder than the other?”

The math teacher leader chimed in to give a specific example of how to do this. She pointed out how two of the Benchmark items assessed students’ knowledge of scientific notation, but in different ways. She followed up by saying that she would work with a small group of students that were having problems with scientific notation at a time that the classroom teachers could observe this as a demonstration lesson.

In this example, the principal pushed teachers towards the standards of the Core Curriculum and raised interesting questions for teacher reflection. The principal and the math teacher leader worked as a tag team; the principal raised a broad point about noticing differences in questions about the same standard and the math leader follows up with specific examples. In this meeting, teachers were expected to bring the Core Curriculum and their Benchmark data and to be prepared to discuss their preliminary analysis of results and what they intended to do.

The principal pushed teachers towards the standards of the Core Curriculum and raised interesting questions for teacher reflection.

In Summary

It is notable that school leaders in all four schools established key organizational structures to support use of the Benchmarks – structures that were not necessarily present in all of the other schools in our sample or across the district. School schedules accommodated regular grade group meetings. In addition, school leaders – the principal and teacher leaders – consistently attended grade group meetings, ensuring that grade teachers actually gathered together and sending a message that the meeting was important. The presence of these leaders provided at least the opportunity for school leaders to learn about teachers’ perspectives on the data, teachers’ understanding of the Core Curriculum, and what instructional strategies teachers were using. Their presence also provided the opportunity for school leaders to signal instructional priorities and draw connections between what was being learned from data in other grades that was relevant to this group of grade teachers. Opportunities for cross-school knowledge were increased, as principals and teacher leaders shared ideas learned in one grade group with others throughout the school. As the examples illustrate, whether and how leaders capitalized on these opportunities varied considerably.

Across the four observations, practitioners used the Item Analysis Report to identify student weaknesses. It is noteworthy that much of the conversation

about remediating gaps focused on a single test item, rather than on curricular standards or instructional approaches that would address these standards. The format of the Item Analysis Report itself may drive practitioners to focus on individual items. This particular report does not group together items testing the same standard and it identifies the standard only by number – thereby requiring that an educator be sitting with the Core Curriculum Standards in order to identify the actual content with which students are struggling. The emphasis on individual items also may contribute to the inordinate amount of time school leaders and teachers spent in discussions about test questions that were poorly worded or otherwise framed in a way that did not make sense or whose content had not been covered in the Core Curriculum yet. In such cases, school leaders need to direct attention back to the curriculum and the standards, as the principal in Example 4 does.

It is important that school leaders have sufficient knowledge about the Benchmarks, the curriculum, and the PSSA so that they can help teachers stay focused on what useful information they can garner from the Benchmarks. For example, understanding the relationship between a fraction and a decimal is one of the “big ideas” in upper elementary mathematics that has the potential to open up a discussion of what is, or is not, in the curriculum for addressing this important concept. The image of an instructional community ready to engage deeply with a content area represents quite a different picture than most discussions about Benchmark data that we observed or heard about.

As a consequence of reviewing Benchmark data, practitioners in the four examples above planned actions that included:

1. **Identifying students who were likely to move from Basic to Proficient or from Below Basic to Basic and targeting them for special interventions in order to increase the likelihood that the school will make AYP.** Across the schools, these interventions varied considerably – extended day programs, Saturday school, work with volunteer tutors, special attention from the math or reading specialist, computer assisted programs. It is likely that their quality varied as well, but formal or informal assessment of the interventions was rare. As one principal told us, “You know, we’ve never really looked to see if those tutors are doing a good job.” (2007)
2. **Identifying skills and concepts to be re-taught in the sixth week of the instructional cycle or in subsequent units.** From our data, we surmise that re-teaching was one of the actions most frequently taken as a result of reviewing the Benchmark results. District leaders and principals reported that there were too many instances of teachers simply returning to the content material, using the same instructional strategies. But some teachers reported that it

It is noteworthy that much of the conversation about remediating gaps focus on a single test item, rather than on curricular standards or instructional approaches.

was important to try different instructional strategies for re-teaching an area of weakness. As one explained,

I can see how my whole class is doing. And they [members of my grade group] can say, "This one question, only four of your twenty kids got it right." So, I know that if only four kids got it right, that's something I need to go back and re-teach, or get a fresh idea about how to give them that information. (Teacher, 2006)

- 3. Identifying students who shared similar weaknesses (or, in some cases, strengths) for re-grouping to provide differentiated instruction.** Our data indicate that re-grouping was another one of the actions most frequently taken as a consequence of reviewing the Benchmark results. Often referred to as "flexible groupings," teachers and school leaders explained that they grouped students around shared weaknesses identified through examination of the Benchmark data. One teacher described how "the groups constantly changed" so that she could "target specific kids and their specific needs and group kids according to where they were lacking." When she felt it was appropriate, she would also assign different homework to different students based on their needs. In other schools, teachers described how they had begun creating groups that cut across classrooms based on shared student weaknesses.

- 4. Re-thinking classroom routines that emphasized greater student independence, motivation, and responsibility for their own learning.** This kind of action was not mentioned frequently. However, one example is a fifth grade teacher who described how she regrouped students, putting stronger students with weaker students as a way to encourage and facilitate peer teaching.

I put the item analysis report on the overhead [for the whole class to see]. It's because of that relationship I have with my students. It's that community. So [I want my students thinking about] why our class average is 60% when I scored 100%. I didn't get any wrong. We need to help our classmate that had difficulty, that may have received 40%. That's where I go into my grouping. How can I pool my strong students [to work with students who are struggling? (May 2007).

- 5. Identifying content and pedagogical needs of teachers to inform opportunities for continued professional learning and other supports that addressed those needs.** Formal professional development sessions and less formal on-the-spot coaching were also planned based on results from the Benchmarks, especially when those data corroborated data from the PSSA. One teacher described a particularly strong approach to supporting teachers' learning:

We actually had a professional development about it, where [the principal] did a lesson to show us, and then we went to two other teachers' rooms and saw them do a lesson. And then pretty much that whole week that followed, [the principal] came around to see how we were using it, if we needed any help, what other support we needed to get this going and into play. (June 2006)

Each of these planned actions makes sense. Each emerged from paying attention to data.” However, the quality of the actions varied considerably. Spillane et al., (2002) argue that educators’ interpretations of policy mandates are critical to their implementation of these mandates.⁴⁶ In the examples above, we note the influence of the accountability environment on educators’ interpretation of the mandate for data-driven decision-making. Clearly, this policy context and the fact that these schools had been identified as “low performing,” influenced practitioners’ perceptions of why examining data is important. They needed to address the primary problem that they felt compelled to solve: how to make AYP. They brought the imperative to “do something” – some might say “do anything” – to their discussion and interpretation of Benchmark data.

However, school leaders can mediate the high stakes accountability environment by creating opportunities for teachers to *learn* from Benchmark data. Beer and Eisenstat (1996) lay out the significance of organized talk to organizational learning:

*Lacking the capacity for open discussion, [practitioners] cannot arrive at a shared diagnosis. Lacking a shared diagnosis, they cannot craft a common vision of the future state or a coherent intervention strategy that successfully negotiates the difficult problems organizational change poses. In short, the low level of competence in most organizations in fashioning an inquiring dialogue inhibits identifying root causes and developing fundamental systemic solutions.*⁴⁷

Our data indicate that the quality of practitioners’ sense-making determines the quality of the actions that they take based on the data. This finding offers insight into why the survey measure – *teacher satisfaction with Benchmarks* – was not a predictor of gains in student achievement. If practitioners focus only on superficial problems – described as “the low-hanging fruit” by principals in our study – their intervention strategies are likely to be mundane.⁴⁸

⁴⁶ Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, 72(3), 387-431.

⁴⁷ Beer, M. & Eisenstat, R. A. (1996). Developing an organization capable of implementing strategy and learning. *Human Relations*, 49(5), 597-619, p. 599-600.

⁴⁸ Sarason, S. B. (1982). *The culture of the school and the problem of change*. Boston: Allyn & Bacon, Inc.

Chapter Five:

Making the Most of Benchmark Data: The Case of Mahoney Elementary School

In this chapter, we use our qualitative data to examine how the multiple factors, that were so difficult to disentangle quantitatively, interact within a school context. While research has emphasized that school leaders are in a position to encourage and support school staff to use data to transform practice,⁴⁹ there remains much to be done in offering detailed examinations of school leaders' work in this area.⁵⁰ Spillane and his colleagues distinguish between "macro functions" (e.g., encouraging data-driven decision-making) and "micro tasks" (e.g., displaying the data, formulating substantive and provocative questions about the data). They urge researchers to analyze how educators "define, present, and carry out these micro tasks" and how the micro-actions interact with one another and with other contextual factors.⁵¹ Our goal was to understand how school leaders build the strong feedback systems that we discussed in Chapter One.

Below, we focus on the Mahoney Elementary School,⁵³ briefly described in Example 4 of Chapter Four. Here, we look in more detail at how school leaders – particularly the principal and subject area teacher leaders – established strong processes for collective learning from Benchmark data within and across instructional communities at Mahoney.⁵² For Mahoney, the Benchmarks were a powerful vehicle for reinforcing the use of the curriculum, for focusing teachers' attention on the standards, and for organizing conversations about student achievement in which teachers were expected to talk about ways to improve their teaching. In effect, these school-based discussions around the Benchmark assessments helped nurture the "*instructional coherence*" cited in Chapter Two and identified by the Consortium for Chicago School Research (CCSR) as showing a positive impact on student learnings.⁵⁴

The Benchmarks were a powerful vehicle for reinforcing the use of the curriculum, for focusing teachers' attention on the standards, and for organizing conversations about student achievement in which teachers were expected to talk about ways to improve their teaching.

⁴⁹ Choppin, J. (2002, April 2). Data use in practice: Examples from the school level. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.; Wohlsetter, P., Datnow, A., & Park, V. (2007, April). Creating a system for data-driven decision-making: Applying the principal - agent framework. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

⁵⁰ Spillane, J. P., Halverson, R. R., & Diamond, J. B. (2001, April). Investigating school leadership practice: A distributed perspective. *Educational Researcher*, 30(3), 23-28.

⁵¹ Spillane, J.P. et al., 2001, p. 24.

⁵² Brown, J. S. & Duguid, P. (2000). Organizational learning and communities of practice: Toward a unified vision of working, learning, and innovation. In Lesser, E. L., Fontaine, M., and Slusher, J. A., *Knowledge and communities* (99-121). Boston: Butterworth Heinemann.; Wenger, E., McDermott, R., & Snyder, W. M. (2002). *Cultivating communities of practice*. Boston: Harvard Business School Press.

⁵³ Pseudonyms are used in this case study for the school and its principal.

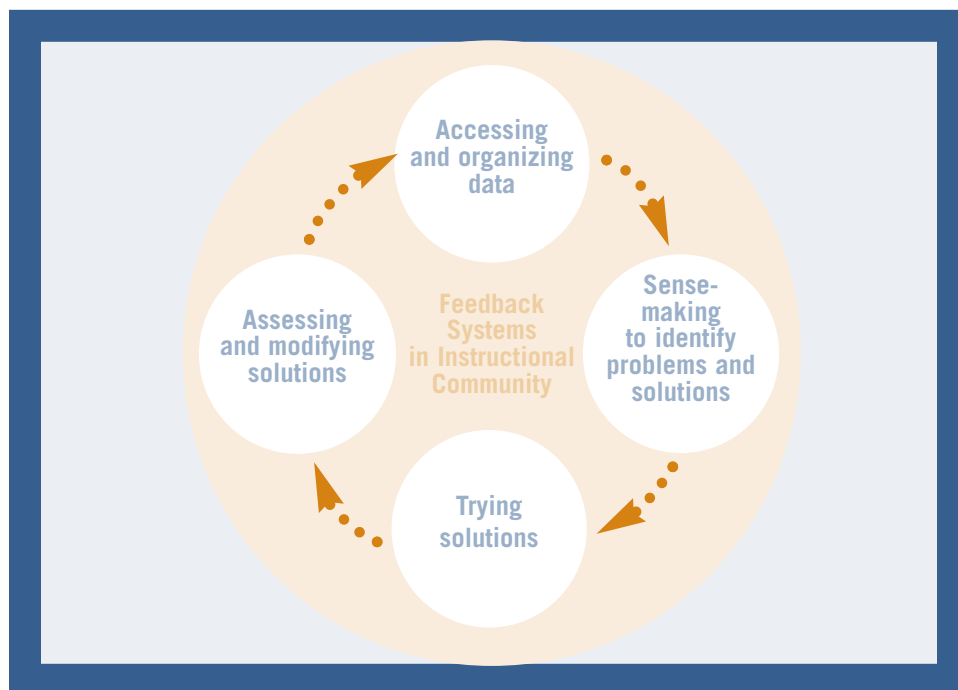
⁵⁴ Newmann, F. M. et al., 2001.

Table 5.1 Interviews and Observations Conducted at Mahoney Elementary School 2005-06 through 2006-07

Researchers conducted intensive fieldwork at Mahoney Elementary School in 2005-06 and 2006-07. During that time, we conducted a total of six observations of leadership team meetings, grade group meetings, CSAP meetings and a school-wide professional development session. We interviewed a total of 11 school staff including the principal, math and literacy leaders, a school secretary and classroom teachers. We interviewed some individuals multiple times.

| Staff Position | 2005-06 Interviews | 2006-07 Interviews |
|--|----------------------|----------------------|
| Principal | 2 | 2 |
| Math leader | 2 | 2 |
| Literacy leader | 1 | 2 |
| Third grade teacher | 1 | |
| Fourth grade teacher A | 1 | 1 |
| Fourth grade teacher B | | 1 |
| Fifth grade teacher A | | 1 |
| Fifth grade teacher B | | 1 |
| Fifth grade teacher C | | 1 |
| Sixth grade teacher | | 1 |
| Secretary | | 1 |
| Setting | 2005-06 Observations | 2006-07 Observations |
| Leadership Team | | 1 |
| Grade Group | 1 | 1 |
| Comprehensive Student Assistance Process | 1 | 1 |
| Professional Development | | 1 |

Figure 5.1 Feedback Loop for Engaging with Data



School Leaders and Effective Feedback Systems

At Mahoney, the principal, Ms. Bannon, established high expectations and a high level of structure to classroom instruction. She participated actively in the school's weekly grade group meetings and worked closely with teacher leaders and classroom teachers to improve instruction. Her high expectations for teachers and students created discomfort for some staff members; however, her commitment to children was respected. Ms. Bannon and the math and literacy teacher leaders orchestrated grade group discussions of Benchmark and other assessment data that built a shared set of goals for teaching and learning and provided an ongoing context for professional learning.

Mahoney's teacher leaders were both fully released from regular classroom instruction. Not only did they work with Ms. Bannon to identify short-term interventions based on Benchmark data at meetings together, they also collaborated with the principal on developing long-term strategies for meeting the school's goals. The principal explained why she had prioritized putting limited resources into full-time teacher leaders when she became the principal a few years before our study began,

"It was a hard decision since it meant larger class sizes. But I wanted to begin with a strong leadership team. It's a choice between having a great teacher reach 25 students or having a great teacher reach other teachers"(2007).

The multiple contributions of the teacher leaders at Mahoney were apparent in both interviews and observations. For example, in our complete fieldnotes for the grade group meeting briefly in Example 4 in Chapter Four, the math teacher leader:

- pointed out that using calculators would improve student scores on a significant number of Benchmark and PSSA (state-wide accountability test) questions;
- offered to conduct a workshop for teachers about how to use their classroom sets of calculators as part of the upcoming professional development day;
- explained that “matrix multiplication” showed up on the Benchmarks, but was a technique that is specific to a particular curriculum and wouldn’t be on the PSSA; and
- provided strategies for teaching the mathematical concept of “expanded notation” and offered to come into the 4th grade classrooms and to model lessons on expanded mathematical notation for small groups of students.

At this meeting the math teacher leader used her knowledge of the Core Curriculum, the Benchmark assessments, and the state’s accountability assessment to help teachers set instructional priorities. She offered suggestions about instructional materials (e.g., calculators). She pointed out the kinds of professional development that the school ought to offer. Perhaps, most importantly, she established why it was important that teachers open their classroom doors and allow her to provide support and guidance through demonstration lessons. Many teachers interviewed, especially in the lower grades, articulated the value of the teacher leaders’ ongoing support. One said, “Knowing that my literacy leader is there [is important], and if I say to her, ‘You know, I’m not really sure how I’m going to do this lesson,’ she’s always there and very helpful.” (2006).

In Chapter One, we posited a four-step feedback cycle as a central element within a school’s overall capacity for data-driven organizational learning and student achievement gains. These steps included school leaders and teachers:

- 1 Accessing and organizing data about students’ understanding of the Core Curriculum (the Benchmark assessments);
- 2 Making sense of the data – both individually and collectively (grade group meetings) – to identify problems and potential solutions;
- 3 Trying the solutions back in their classrooms; and
- 4 Assessing and modifying their solutions based on classroom assessments.

As discussed in Chapter Two, the school district intended for the Benchmark assessments to provide the kind of formative feedback that allows teachers to make mid-course corrections in their instructional strategies. Teacher

“[Allocating the resources for full-time content area teacher leaders] was a hard decision since it meant larger class sizes. But I wanted to begin with a strong leadership team. It’s a choice between having a great teacher reach 25 students or having a great teacher reach other teachers.”

- Mahoney Principal

leaders at Mahoney were critical to the school's success in implementing systems and an organizational culture that enabled these kinds of feedback systems across the school. In any cycle, the "linkages" that connect the steps are crucial and are often the weak points in a system (See [Figure 5.1](#)). Teacher leaders helped support those links, and in many cases served as links themselves, sharing knowledge from grade group meetings across the school.

Additionally, review of Benchmark data at Mahoney was integrated into the kinds of *feedback systems* discussed in Chapter One. Teachers experimented with new practices that had been identified in grade group meetings. School leaders followed up in classrooms to help teachers with new instructional strategies and to modify these practices where appropriate. These steps became routine at Mahoney, thus ensuring that feedback systems were strong and coherent during the period of our research.

"Knowing that my literacy leader is there [is important], and if I say to her, 'You know, I'm not really sure how I'm going to do this lesson,' she's always there and very helpful."

- Mahoney Teacher

Grade Group Meetings and Benchmark Discussions

Grade group meetings were a key opportunity for looking at and learning from Benchmark data at Mahoney. These meetings were held weekly and included the principal, the math teacher leader, the literacy teacher leader, and the two or three classroom teachers for each grade. Grade group meetings were described by the principal and teacher leaders as *the* most important site in the school for teacher learning. In fact, during the second year of our research, Ms. Bannon reported that they had decided to call the meetings "Professional Learning Communities" instead of grade groups, to highlight their contribution to teachers' professional learning.

Grade group meetings at Mahoney were highly structured and consistently focused on instructional issues. Each meeting began with a member of the leadership team handing out a typed agenda with a guiding question at the top, ended with the principal summarizing next steps, and was followed up with typed notes distributed to all participants. According to teachers and school leaders, grade group meetings always focused on analysis of data or reflection on instruction. As one teacher told us, "Everything begins by talking about data."

The Benchmark Item Analysis Reports were important tools in grade group meetings, as they were in other schools. At Mahoney, however, the Core Curriculum Standards was another key tool in grade group meetings. Teachers were expected to bring the curriculum framework to grade group meetings so they could refer to it as they discussed the standards in which their students showed weaknesses. In addition, teachers were expected to prepare for grade group meetings by filling out the district's Benchmark

Data Analysis Protocol, which asked them to assess students' weaknesses and identify strategies for improving the areas of weakness. They used these protocols in conversations with their colleagues. The structure of the meetings themselves supported the continuity of the feedback system. Use of the same formats and reports created a common framework and language. Clear follow-up about next steps ensured that the momentum of the meeting was not lost.

The heart of the grade group meetings was the discussion of Benchmark and other assessment data. As in other schools, Mahoney's grade group discussions of Benchmarks encompassed what we identified earlier in Chapter Four as three interconnected types of sense-making: *strategic* (e.g., short-term tactics to help the school reach AYP), *affective* (teachers' beliefs about their students and their collective responsibility for student learning), and *reflective* (evaluating their own instructional practices and connecting Benchmark data to with key curriculum concepts).

Analysis and discussion of Benchmark data not only focused on instruction, but also highlighted the interim assessments' connection to other accountability tests, an example of strategic sense making. Teachers and leaders discussed how many and which students were close to Proficient or Advanced – performance categories on the PSSA. Talk about Benchmarks and the PSSA also led to talk about the school's moral purpose and the leaders' belief in the capabilities of their staff and students. In one grade group meeting, Ms. Bannon commented that the cut-off points for identifying individual students as Advanced and Proficient were too low, saying that “we have to set our own goal as higher than that”(2005). The expectation that all students would be Proficient was accompanied by a consistent focus in grade group meetings on the Core Curriculum, the standards, and what teachers could do to improve their own teaching. As one teacher said:

The school has been focused on using the data to help the kids and push the instruction. Every kind of thing that we do, every assessment we give, we look at it; we see what we need to change, and how we can differentiate our instruction so that it's helping them do more. (2006)

Teachers at Mahoney were pushed to question their own past practices and they both sought and shared new ways to approach content that needed to be taught and new ways to help their students learn. The re-naming of the grade group meetings as “Professional Learning Communities” was appropriate.

[The principal] commented that the cut-off points for identifying individual students as Advanced and Proficient were too low, saying that “we have to set our own goal as higher than that.”

Organizational Learning and Instructional Coherence

In summary, the principal and teachers leaders at Mahoney had a clear understanding of the powerful connection between the Benchmarks and the Core Curriculum and their importance to establishing instructional coherence across the school. The principal allocated resources for knowledgeable teacher leaders who were expert in the content and assessment issues in their own curricular areas. Together, the principal and teacher leaders established a set of structures and practices that ensured that Benchmark data were used as part of a process for ensuring high quality instruction within and across grade groups, as well as other settings in the school. At Mahoney, the principal and the teacher leaders were “learning leaders,” who created a climate in which adult learning was central to school improvement.⁵⁵ They took the lead in helping teachers sift through reams of data and make sense of competing priorities. Leadership around the use Benchmark data was *distributed* across the roles of principal and teacher leaders.⁵⁶ Alongside principals, teacher leaders can assume important leadership functions relative to data use.

Alongside principals, teacher leaders can assume important leadership functions relative to data use.

⁵⁵ Elmore, R. F. (2000, December). *Building a new structure for school leadership*. Washington, DC: The Albert Shanker Institute.; DuFour, R. (2002, May). The learning-centered principal. *Educational Leadership*, 59(8), 12-15.; Spiri, M. H. (2001, May). *School leadership and reform: Case studies of Philadelphia principals*. Philadelphia, PA: Consortium for Policy Research in Education.

⁵⁶ Spillane, J.P. et al., 2001.

Making the Most of Benchmark Data at Mahoney Elementary School

Engaged Principal:

- Built strong leadership team by allocating full time teacher leaders in math and reading
- Worked with teacher leaders to develop long-term instructional improvement strategies and shorter-term priorities for their work with classroom teachers
- Emphasized data-driven decision-making
- Actively attended grade group meetings
- Established meeting routines that were used across the school
- Set high expectations for teachers' preparation for and participation in grade group meetings
- Used discussions of Benchmark data in grade groups to reinforce importance of proficiency standards of Core Curriculum
- Encouraged strategic, affective, and reflective sense-making, with the strongest emphasis on reflective sense-making
- Worked with teacher leaders to spread insights and knowledge about instruction across the school

Full-time Math and Reading Teacher Leaders:

- Well-versed in the Core Curriculum, the Benchmark assessments, and the PSSA exams and understood the connections and disconnections among the three
- Continuously enhanced their knowledge of research-based instructional strategies that supported effective use of the Core Curriculum
- Helped teachers interpret Benchmark data
- Recommended specific instructional strategies based on the Benchmark results
- Moved in and out of classrooms to see if teachers were implementing curriculum well and provided coaching and demonstration where needed
- Gathered resources to supplement the curriculum
- Collaborated with principal on long and shorter-term instructional strategies to meet school's goals

Effective Grade Group Meetings:

- Held weekly
- Principal, teacher leaders, and classroom teachers came prepared to participate
- Discussions included strategic, affective, and reflective sense-making
- Highly structured meeting routines, focused on instructional issues and ongoing professional learning of staff
- Began with an agenda and guiding question
- Ended with school leader summarizing next steps
- Follow-up notes distributed across the school

Conclusion

Making the Most of Interim Assessment Data: Implications for Philadelphia and Beyond

Federal, state, and district policies that use standardized tests as the central metric for accountability have fueled the fervor for student achievement data, especially in districts with large numbers of academically failing students. The rise of interim assessments is inextricably tied to the policy environment of No Child Left Behind. Controversy notwithstanding, the use of interim assessments by large urban school districts to improve instruction and student achievement is on the rise. The findings from our research on the use and impact of these assessments in Philadelphia's K-8 schools will not end the debate. They do, however, offer formative lessons to Philadelphia and beyond about the design, implementation, and impact of interim assessments. Below, we discuss the implications of this research for policy makers and district and school leaders. The research also has important implications for the higher education community that educates and certifies district and school leaders.

Data can make problems more visible, but only people can solve them.

Investing in School Leaders

The most important message from this research is that the success of even a well-designed system of interim assessments is dependent on the knowledge and skills of the school leaders and teachers who are responsible for bringing the system to life in schools. Stringent accountability measures, strong curricular guidance, and periodic assessments are not substitutes for skilled and knowledgeable practitioners. Data can make problems more visible, but only people can solve them.

In addition, mandated accountability measures, in and of themselves, are an inadequate foundation for building the kinds of collegial relationships that result in shared responsibility for school improvement and improved student learning.

In Philadelphia, the very federal and state policies that persuaded district leaders and school practitioners to pay careful attention to data, also constrained their ability to make the most of Benchmark results for improving instruction and student achievement. Immediate needs for improved testing outcomes often worked against practitioners learning more about how to help all students master the concepts and skills of the Core Curriculum.

However, our research also indicates that the use of Benchmark data is not always a narrow exercise in preparing to “teach to the test.” We witnessed how school leaders were able to mediate the often counter-productive environment of high stakes accountability. In the language of organizational learning, these leaders enacted organizational practices that contributed to individual teacher learning and professional growth, while at the same time fortifying a collective understanding of the challenges, goals, and path ahead for the school.

Data-driven decision-making represents a new way of thinking for most educators. And, as this report has demonstrated, the logic of data use is built on numerous assumptions that cannot be taken for granted, especially the ability of school leaders to help teachers make the most of Benchmark results. Organizational learning offers a robust framework for understanding what school leaders need to know and be able to do in order to make the most of interim assessment results and other kinds of data about student achievement.

School leaders need to be able to lead the kinds of deliberative conversations that create opportunities for teacher learning.

- **As learning leaders, principals and teacher leaders need to know how to facilitate “learning” discussions about data.** School leaders can make a real difference in helping staff move beyond data use as a narrow exercise in preparing to “teach to the test.” But to do so, they must know how to frame conversations about assessment data so that teachers understand the connections to larger school improvement priorities and to the curriculum. They need to know how to pose questions that invite teachers to talk openly about: curriculum concepts, how their students learn best, what instructional practices have worked and those that haven’t, what additional curricular resources they need, what they need to learn about content, and where they might seek evidence-based instructional strategies that would address the learning weaknesses of their students. They also need to be able to steer teachers away from inappropriate use of Benchmark data, such as predicting performance on the PSSA. School leaders need opportunities to practice these skills and receive feedback. Understanding the value and purposes of the different types of sense-making identified in our research – affective, strategic, and reflective – and how to use them offer a framework for such training.

- **As learning leaders, principals and teacher leaders need to know how to allocate resources and establish school organizational structures and routines that support the work of instructional communities and assure that the use of Benchmark data is embedded in the feedback systems necessary for organizational learning.** School schedules need to accommodate regular meetings of grade groups. Principals and teacher leaders need to be at these meetings and, with teachers, establish meeting routines that include agendas, discussion protocols with guiding questions, and documentation of proceedings. Follow up to the meetings is crucial. School leaders need to visit classrooms to see if and how teachers are using instructional strategies and to offer resources and coaching so that teachers can deepen their understanding of curriculum content and pedagogy. Assessing the impact of interventions is also crucial. Important steps include helping teachers to design classroom based assessments for use during the sixth week of instruction and examining the quality of common interventions such as tutoring and after-school remediation programs. School leaders must recognize their role in the creation and diffusion of knowledge across the school.

Designing Interim Assessments and Supports for Their Use

This research also offers lessons about designing interim assessments and the resources that will encourage and support the use of data from those assessments. Philadelphia's Benchmark assessments have a number of clear design strengths that may offer guidance to other districts considering adoption of interim assessments. The alignment of the Benchmarks with the Core Curriculum reinforced expectations for what teachers should teach and at what pace; it made the Benchmark results highly relevant to teachers' instructional planning. The timely return of the results and the allocation of a sixth week for re-teaching after review of the data buttressed the instructional intention of the Benchmarks. District supports in the form of technology, tools for data analysis and interpretation, and professional development were largely appreciated by school staff. All of these elements likely contributed to broad acceptance and use of the Core Curriculum and Benchmark assessments by Philadelphia K-8 teachers.

- **As districts and schools develop organizational structures, processes and tools to support the use of interim assessment data, they need to ask themselves these questions:**

Do the structures, processes, and tools support the review of data as a collective learning activity of instructional communities? Are they supporting the review of data as an activity which helps teachers deepen their pedagogical content knowledge and understand what their students know and how they learn?

Do they support the multiple steps of feedback loops? Do they encourage leaders' follow-up work with teachers in classroom? Do they promote the assessment of interventions and modifications where necessary?

- **In Philadelphia, district leaders should revisit their purposes for the Benchmark assessments with the goal of prioritizing one or two purposes. To achieve the instructional purposes that district leaders intended, it is likely that the Benchmark assessments are in need of modifications.**

In order to capitalize on Benchmarks to fulfill instructional purposes, the district leaders should: review Benchmark items to make certain that they: test for a range of thinking skills – knowledge, comprehension, application, synthesis and evaluation – and that they offer distractor answers that provide insight into what students don't understand. Continued efforts should be made by the district and testing industry to include open-ended items.

Implications for Future Research

We believe that the use of a multi-method design and organizational learning as an analytic framework were two strengths of this study. Used in concert, they offer considerable promise in unraveling the connections among many factors related to the use of data in schools and gains in student achievement. There are numerous refinements to our approach that researchers might make that would make significant contributions to both theory and practice. These include more direct survey measures of data use and analyses at the classroom and instructional community levels.

We also realize that we only scratched the surface in terms of the three kinds of sense-making and the relationships between the kinds of sense-making and the resulting instructional plans. We suggest that discourse analysis offers a robust methodology for research on data use and instructional improvement.

One of the controversies surrounding interim assessments is whether they actually serve formative purposes for teachers and students. While we, as well as other researchers, have begun to build a knowledge base about the impact of interim assessments on *teachers'* instructional practice, there remains much work to do on whether interim assessment results help *students* understand their mistakes and make appropriate adjustments in their thinking.

Reference List

- A curriculum audit of the Philadelphia public schools, Philadelphia, PA. International Curriculum Management Audit Center. *Phi Delta Kappa, International*. May 16-21, 2005.
- Argyris, C. & Schon, D. A. (1978). *Organizational learning: A theory of action perspective*. Reading, MA: Addison-Wesley.
- Beer, M. & Eisenstat, R. A. (1996). Developing an organization capable of implementing strategy and learning. *Human Relations*, 49(5), 597-619, p. 599-600.
- Black, P. & Wiliam, D. (1998, October). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*.
- Blume, H. (2009, January 28). L.A. teachers' union calls for boycott of testing. *Los Angeles Times* [On-line]. Retrieved on February 11, 2009 from <http://www.latimes.com/news/education/la-me-laUSD28-2009jan28,0,4533508.story>.
- Brown, J. S. & Duguid, P. (1998). Organizing knowledge. *California Management Review*, 40(3), 28-44, p. 28.
- Brown, J. S. & Duguid, P. (2000). Organizational learning and communities of practice: Toward a unified vision of working, learning, and innovation. In Lesser, E. L., Fontaine, M., and Slusher, J. A., *Knowledge and Communities* (99-121). Butterworth Heinemann.
- Bulkley, K. E., Mundell, L., & Riffer, M. (2004). *Contracting out schools: The first year of the Philadelphia Diverse Provider Model*. Philadelphia: Research for Action.
- Burch, P. (2005, December 15). The new education privatization: Educational contracting and high stakes accountability. *Teachers College Record*.
- Cech, S. J. (2008, September 17). Test industry split over 'formative' assessments. *Education Week*, 28(4), 1, 15, p. 1.
- Century, J. R. (2000). Capacity. In N. L. Webb, J. R. Century, N. Davila, D. Heck, & E. Osthoff (Eds.), *Evaluation of systemic change in mathematics and science education*. Unpublished manuscript, University of Wisconsin-Madison, Wisconsin Center for Education Research.
- Choppin, J. (2002, April 2). Data use in practice: Examples from the school level. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Clune, W. H. & White, P. A. (2008, October). *Policy effectiveness of interim assessments in Providence Public Schools*. WCER Working Paper No. 2008-10. Wisconsin Center for Education Research, School of Education, University of Wisconsin-Madison <http://www.wcer.wisc.edu/>. p. 5.
- Corcoran, T. B. & Christman, J. B. (2002, November). *The limits and contradictions of systemic reform: The Philadelphia story*. Philadelphia: Consortium for Policy Research in Education.
- Daft, R. L. & Weick, K. E. (1984). Towards a model of organizations as interpretation systems. *Academy of Management Review*, 9(2), 284-295.
- DuFour, R. (2002, May). The learning-centered principal. *Educational Leadership*, 59(8), 12-15.

Elmore, R. F. (2000, December). *Building a new structure for school leadership*. Washington, DC: The Albert Shanker Institute.

Halverson, R. R., Prichett, R. B., & Watson, J. G. (2007). *Formative feedback systems and the new instructional leadership* (WCER Working Paper No. 2007-3). [On-line]. Retrieved on July 16, 2007, from <http://www.wcer.wisc.edu/publications/workingPapers/index.php>.

Knapp, M. S. (1997). Between systemic reforms and the mathematics and science classroom: The dynamics of innovation, implementation, and professional learning. *Review of Educational Research*, 67(2), 227-266.

Leithwood, K., Aitken, R., & Jantzi, D. (2001). *Making schools smarter: A system for monitoring school and district progress*. Thousand Oaks, CA: Corwin Press.

Lipsey, M.W., and D.B. Wilson (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48, 1181-1209.

Little, J. W. (1999). Teachers' professional development in the context of high school reform: Findings from a three-year study of restructuring high schools. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Quebec.

Mason, S. A. & Watson, J. G. (2003). Understanding schools' capacity to use data. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Mintz, E., Fiarman, S. E., & Buffett, T. (2006). Digging into data. In K. P. Boudett, E. A. City, & R. J. Murname (Eds.), *Data wise: A step-by-step guide to using assessment results to improve teaching and learning* (81-96). Cambridge, MA: *Harvard Education Press*, p. 94.

Newmann, F. M., Smith, B., Allensworth, E., & Bryk, A. S. (2001, January). *Improving Chicago's schools: School instructional program coherence benefits and challenges*. Chicago: Consortium on Chicago School Research.

Newmann, F. M., Smith, B., Allensworth, E., & Bryk, A. S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis*, 23, 297-321.

Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007, November). *The role of interim assessments in a comprehensive assessment system*. Washington, DC: The Aspen Institute.

Porter, A. C., Chester, M. D., & Schlesinger, M. D. (2004, June). Framework for an effective assessment and accountability program: The Philadelphia example. *Teachers College Record*, 106(6), 1358-1400.

Resnick, L. B. & Hall, M. W. (1998). Learning organizations for sustainable education reform. *Journal of the American Academy of Arts and Sciences*, 127(4), 89-118, p. 108.

Rusch, E. A. (2005). Institutional barriers to organizational learning in school systems: The power of silence. *Educational Administration Quarterly*, 41, 83 – 120. [On-line]. Retrieved on May 8, 2007, from SAGE Full-Text Collections.

Sarason, S. B. (1982). *The culture of the school and the problem of change*. Boston: Allyn & Bacon, Inc.

Senge, P. (1990). *The fifth discipline: The art & practice of the learning organization*. NY: Doubleday.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1-22.

Spillane, J. P., Halverson, R. R., & Diamond, J. B. (2001, April). Investigating school leadership practice: A distributed perspective. *Educational Researcher*, 30(3), 23-28.

Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, 72(3), 387-431.

Spillane, J. P. & Thompson, C. L. (1997, June). Reconstructing conceptions of local capacity: The local education agency's capacity for ambitious instructional reform. *Education Evaluation and Policy Analysis*, 19(2), 185-203.

Spillane, J. P. & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. *Educational Evaluation and Policy Analysis*, 21(1), 1-27.

Spiri, M. H. (2001, May). *School leadership and reform: Case studies of Philadelphia principals*. Philadelphia, PA: Consortium for Policy Research in Education.

Travers, E. (2003, September). Philadelphia school reform: Historical roots and reflections on the 2002-2003 school year under state takeover. *Penn GSE Perspectives on Urban Education*, 2(2).

Useem, E. (2005, August). Learning from Philadelphia's school reform: What do the research findings show so far? Paper presented at the No Child Left Behind Conference, Sociology of Education Section of the American Sociological Association, Philadelphia, PA.

Wagner, T. (1998). Change as collaborative inquiry: A 'constructivist' methodology for reinventing schools. *Phi Delta Kappan*, 80(7), 378-383.

Wenger, E., McDermott, R., & Snyder, W. M. (2002). *Cultivating communities of practice*. Boston: Harvard Business School Press.

Wohlsetter, P., Datnow, A., & Park, V. (2007, April). Creating a system for data-driven decision-making: Applying the principal-agent framework. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Appendix A Phase One Qualitative Research — School Characteristics

2006-07 Data

| School | Provider | Grades | Number of Students | % from Low-Income Families | Racial/Ethnic Make-up | Achievement % Advanced & Proficient Reading/Math |
|-------------|--------------------------------|--------|--------------------|----------------------------|---|--|
| Lea* | University of Pennsylvania | K-8 | 425 | 85.7 | 91.1% African American .05% White 5.4% Asian 1.4% Latino 1.6% Other | 5th Grade 27.3/42.0 8th Grade 42.6/52.7 |
| Anderson | Edison Schools, Inc. | K-5 | 465 | 80.8 | 97.8% African American 1.3% White 0.6% Latino 0.2% Other | 5th Grade 22.4/46.9 |
| Wright | Victory Schools, Inc. | K-6 | 390 | 86.5 | 97.9% African American 1.0% White 0.5% Latino 0.5% Other | 5th Grade 17.8/20.0 |
| McKinley | Office of Restructured Schools | K-8 | 399 | 86.9 | 23.1% African American 1.0% White 75.4% Latino 0.5% Other | 5th Grade 28.6/60.0 8th Grade 68.1/44.7 |
| EM Stanton | Universal Companies | K-6 | 193 | 85.7 | 93.3% African American 1.6% White 4.7% Latino 0.5% Asian | 5th Grade 70.0/75.0 |
| MH Stanton* | Office of Restructured Schools | K-6 | 412 | 90.4 | 98.1% African American 0.2% White 1.7% Latino | 5th Grade 49.2/77.1 |
| Cooke* | "Sweet 16" | K-8 | 635 | 85.4 | 83.9% African American 0.5% White 8.2% Latino 7.2% Asian 0.2% Other | 5th Grade 8.7/30.4 8th Grade 43.8/35.3 |
| Fulton* | Foundations, Inc. | K-6 | 391 | 90.0 | 95.4% African American 2.0% White 2.3% Latino 0.3% Other | 5th Grade 4.3/27.7 |
| Ludlow* | Edison Schools, Inc. | K-8 | 311 | 86.3 | 59.8% African American 1.3% White 36.0% Latino 2.9% Other | 5th Grade 14.7/38.3 8th Grade 27.3/39.4 |
| Meade | Temple University | K-8 | 463 | 91.7 | 99.4% African American 0.6% Latino | 5th Grade 29.5/37.8 8th Grade 41.7/36.2 |

* Case Study schools 2006-2007.

Appendix B Benchmark Item Analysis Form

| Standards Mastery | Item Analysis | Item Analysis | Student List | Student Analysis | | | | | |
|---|--|---------------|--------------|------------------|---------|----------|----------|---------|----------|
| Subject: | All | | | | | | | | |
| Grade: | 8 | | | | | | | | |
| Test: | 09/27/04 - MAR2004 CCAT Grade 8 Math Test Detail | | | | | | | | |
| Test Name: | MAR2004 CCAT Grade 8 Math | | | | | | | | |
| Test Administration Date: | Monday, September 27, 2004 | | | | | | | | |
| Number of Questions: | 20 | | | | | | | | |
| Total Possible Points: | 20 | | | | | | | | |
| Section: | Homeroom: 8th Grade - (B7) | | | | | | | | |
| Teacher: | MAC RYBARCZYK | | | | | | | | |
| Number of Students: | 34 | | | | | | | | |
| Number of Students Tested: | 30 | | | | | | | | |
| Key: = Correct Response RED LETTER = Actual Incorrect Response Black Digit = Points Scored | | | | | | | | | |
| Question | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Standard | 2.2.8.B | 2.2.8.B | 2.2.8.B | 2.2.8.B | 2.2.8.B | 2.9.8.B | 2.9.8.F | 2.9.8.F | 2.9.8.G |
| Possible Points | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Correct Response | C | C | C | B | B | A | D | A | B |
| BOURNIGHT, KANISHA | | B | | D | | C | | | |
| BURGESON, ALVA | | | | | | C | B | | |
| CHRISTLEY, ANTWAN | | A | | | | B | B | | C |
| DAENALL, DILLON | | | | | | | | | |
| DELANGEL, SCOTT | | | | A | | | | | |
| DOEPKE, LELAH | | | | | | | B | | |
| EASTMAN, ISAH | | | | | | | | | |
| ESSEX, WILLIS | | | | | | | B | | C |
| FALLEN, ZITA | | | | | | C | | | |
| GOLSON, RACAWN | | | | | | D | | | D |
| HECROCK, JAMEL | | | | | | D | | | |

Appendix C List of Topics Covered in Interviews

The following are lists of topics covered in interviews with principals, teacher leaders and classroom teachers. Each round of interviews (Fall 2005, Spring 2006, Fall 2006 and Spring 2007) covered a different, though sometimes overlapping, set of topics.

2005-06 Interview Topics

- School context
 - School's history with reform
 - Current reform initiatives
 - Principal's leadership style
- Changes in and rationale for instructional priorities
 - Identify and explain classroom changes and previous practices
 - Staff and other influences that led to instructional changes
 - Resources necessary for instructional changes
- Leadership team and other instructional communities (grade groups, SLCs)
 - Composition of the leadership team and instructional communities
 - Members' roles, settings for meetings
 - Relationships with the provider
 - Examples of instructional decisions and use of data
- Roles and responsibilities around data
 - Principal's and leadership team's role in using data
 - Provider's role and expectations
 - Responsibilities around organizing and analyzing the data
- Benchmarks and other formative assessment
 - Importance and use of formative assessments
 - Provider and others' role in using formative assessments
- Professional development about data
 - Settings and topics of professional development sessions
- Staff capacity for data
 - Examples of sophisticated and unsophisticated data use
- Resources necessary to use data effectively
 - Technology
 - Human support
- Professional development around data use
- Data analysis tools
 - Identify and describe data analysis tools
 - People and processes involved in implementing the tools
- Useful/helpful data
 - Data used to inform classroom instruction or identify broad problems
 - How were benchmarks used?
 - Useful tools and formats for data analysis
- Settings for discussions and analysis of data

2006-07 Interview Topics

- Context surrounding school leadership
 - Leadership styles and influences on classroom instruction
 - Leadership actions that have influenced instruction
 - Background and self-assessment of effectiveness in school role
 - Sources of support and guidance for teachers and leaders
 - Thoughts on leading in a high stakes environment
 - Role of formal and informal teacher leadership
- School Improvement Planning (SIP)
 - Progress on improvement goals and future priorities
 - Process for planning the goals and priorities
- Instructional changes
 - Changes that school leaders have encouraged and the role of data in promoting those changes
 - Instructional communities and grade groups
 - Structure and roles of the groups
 - Groups' roles in encouraging and guiding teachers,
 - Challenges the groups face
- Data use
 - Instructional changes made because of data
 - Data that teachers have used and found helpful
 - Settings for examining data
 - Tools teachers used to examine data
 - Benchmarks and PSSA writing rubric
 - Where and when do teachers use these tools?
 - What do they learn from each kind of assessment?
- Professional development
 - Types of professional development
 - Impact of the professional development
 - School leaders' roles in professional development sessions
- Impact of high stakes accountability environment
 - Guidance and support from colleagues and leaders

APPENDIX D Technical Details on Data and Methods

Survey Data

The teacher survey was distributed through the schools, and completed surveys were collected and returned by the schools to the district's research office. The survey did not ask teachers to provide their names or other information that could identify them as individuals. Still, some teachers, especially those who work in schools where social trust is low, are wary of completing surveys. It is also notoriously difficult to compel a busy teacher to complete a long survey, which, in this case, involved hundreds of questions spread over 16 pages. Given these challenges, the response rates for the surveys are respectable. A total of 6,680 teachers (65 percent of all teachers) from 204 of 280 schools responded to the spring 2006 survey. A total of 6,007 teachers (60 percent of all teachers) responded to the spring 2007 survey. These response rates are comparable to that for large-scale teacher surveys in other major cities; for example, teacher surveys fielded by the Consortium on Chicago School Research typically produce a response rate of about 60 percent.

To make the school-level predictor variables used in the multilevel models, data from all teachers who responded to the survey (not just teachers in Benchmarks grades and subjects) was aggregated. Schools at which fewer than 30 percent of the teachers responded were excluded from the analysis.

Assessment of Student Learning: The Rank-Based Z-Score Method

During the school years 2004-2005, 2005-2006, and 2006-2007, Philadelphia students in grades three through eight took standardized tests of achievement in reading and mathematics at the end of the school year. However, in some grades, students took the Terra Nova test, a commercially available assessment developed by CTB McGraw Hill. In other grades, students took an assessment developed by Commonwealth of Pennsylvania (PSSA). The different assessments taken in different years necessitate a special strategy to examine learning gains.

To create a comparable indicator of achievement, we placed student scores on the rank-based z-score scale. The rank-based z-score converts a student's percentile (in the Philadelphia distribution of scores) to their position in the normal distribution, so a student at the 50th percentile would have a rank-based z-score of 0, while one at the 95th percentile would have a rank-based z-score of 1.64, and one at the 5th percentile would have a score of -1.64. The indicator of learning growth was created by subtracting the z-score at the end of Year 1 from the z-score at the end of Year 2.

This method is the same used by RAND in its recent reports on the impact on student achievement of privatization of schools in Philadelphia (Gill, Zimmer, Christman, & Blanc, 2007) and on Philadelphia's charter schools (Zimmer, Blanc, Gill, & Christman, 2008).

Technical Description of the Multilevel Models

The dependent variable was the student's rank-based z-score on reading comprehension or mathematics at Time 2 (that is, either the score from spring 2006 or spring 2007). The equations are as follows:

Level 1

$$Y_{ij} = \quad 0_j + \quad 1_j(\text{Race/Ethnicity})_{ij} + \quad 2_j(\text{Gender})_{ij} + \quad 3_j(\text{Special Education})_{ij} + \quad 4_j(\text{Grade at Test 1})_{ij} + \quad 5_j(\text{Grade at Test 2})_{ij} \\ + \quad 6_j(\text{Rank-based z-score on Test at Time 1})_{ij} + r_{ij}$$

Level 2

$$0_j = \quad 00 + \quad 01(\text{Percent Low Income})_j + \quad 02(\text{Additional School-Level Variables})_j + u_{0j}$$

All predictor variables were grand-mean centered.

Appendix E Technical Detail on Scales Used in Chapter 3

The first four scales presented here – Instructional Leadership, Teacher-Teacher Trust, Instructional Innovation and Improvement, and Teacher Collective Responsibility – incorporate *most* of the specific items that make up the indicators with those names developed by the Consortium on Chicago School Research (CCSR). Information on the CCSR scales can be accessed at <http://ccsr.uchicago.edu/content/page.php?cat=4>. The specific items that comprise the scales used in this chapter are shown below. Likewise, the values for Cronbach's alpha were created for these scales from the Philadelphia teacher survey data.

Instructional Leadership

(Eight items; Cronbach's alpha: .94)

To what extent do you disagree or agree with the following statements?

(Response categories: Strongly Disagree, Disagree, Agree, Strongly Agree)

The leadership at this school:

- Makes clear to the staff the expectations for meeting instructional goals.
- Communicates a clear vision for our school.
- Sets high standards for student learning.
- Carefully tracks student academic progress.
- Encourages teachers to implement what they have learned in professional development.
- Knows what's going on in my classroom.
- Actively monitors the quality of teaching in this school.
- Has made data-driven decision-making a priority at the school.

Teacher Commitment to the School

(Four items; Cronbach's alpha: .84)

To what extent do you disagree or agree with the following statements?

(Response categories: Strongly Disagree, Disagree, Agree, Strongly Agree)

- I usually look forward to each working day at this school.
- I wouldn't want to work in any other school.
- I would recommend this school to parents seeking a place for their child.
- Teachers at this school respect other colleagues who are experts at their craft.

Instructional Innovation and Improvement

(Three items; Cronbach's alpha: .90)

How many teachers in this school:

(Response categories: None, Some, About Half, Most, All)

- Set high standards for themselves?
- Are willing to try new ideas?
- Are really trying to improve their teaching?

Teacher Collective Responsibility

(Four items; Cronbach's alpha: .86)

How many teachers in this school:

(Response categories: Some, About Half, Most, All, None)

- Help maintain discipline in the entire school, not just their classroom?
- Take responsibility for improving the school?
- Feel responsible for helping each other do their best?
- Feel responsible when students in this school fail?

Use of the Core Curriculum (Spring 2006)

(Three items; Cronbach's alpha: .89)

I use the Core Curriculum:

(Response categories: Never, Occasionally, Often, Always)

- To guide subject/topic coverage
- To organize and develop instructional units and classroom activities
- To redesign assessment strategies

Use of the Core Curriculum (Spring 2007)

(Four items; Cronbach's alpha: .89)

During the past twelve months, how often did you use the following components of the District's Core Curriculum?

(Response categories: Never, Occasionally, Often, Always)

- The Planning and Scheduling Timeline
- The Writing Plan
- The Course of Study and Prerequisite Skills
- The Coordinating Documents

Usefulness of Benchmarks to Inform Instruction

(Seven items; Cronbach's Alpha:.92)

To what extent do you disagree or agree with the following questions?

(Response categories: Strongly Agree, Agree, Disagree, Strongly Disagree)

- Benchmark test scores give me information about my students that I didn't already know.
- The Benchmarks set an appropriate pace for teaching the curriculum to my students.
- Results on the Benchmark tests give me a good indication of what students are learning in my classroom.
- At my school, the use of Benchmark tests has improved instruction for students with skill gaps.
- The Benchmark tests are a useful tool for identifying the content descriptors that students do and do not understand.
- The Benchmark tests are a useful tool for identifying students' misunderstandings and errors in their reasoning.
- The Benchmark tests are a useful tool for helping students identify what they know and what they still need to learn.

Collective Examination of Benchmarks

(Three items; Cronbach's alpha: .86)

- During the past 12 months, how often did the following occur in your school?
(Response categories: Never, 1-2 times, 3-5 times, More than 5 times)
- Your grade group, field coordinators, or coaches met to discuss ideas for re-teaching a skill that students were lacking, according to the Benchmark test.
- Your grade group, field coordinators, or coaches met to discuss re-grouping students for instruction on the basis of Benchmarks scores.

Access to and Support for Technology Use

(Four items; Cronbach's alpha: .76)

Does the following exist in your classroom or school?

(Response categories: Yes, No)

- Internet in the classroom

To what extent do you disagree or agree with the following statements?

(Response categories: Strongly Disagree, Disagree, Agree, Strongly Agree)

- Our school's technology coordinator helps teachers integrate computing technology into lessons.
- I can find help in my school when I have trouble using computing technology.
- The computing technology in my school is in good working order.

Professional Development on Using Data

(Four items; Cronbach's Alpha: .84)

Over the past 12 months, which of the following have been the focus of a professional development session, faculty meeting, grade group meeting, or subject area meeting?

(Response categories: Check all that apply)

- Accessing your students' performance data on the computer
- Principal and/or school leadership team presentation about your school's performance data
- Using student performance data to develop an action plan
- Using student performance data to assess the effectiveness of teaching practice

Authors

Jolley Bruce Christman

Jolley Bruce Christman, Ph.D. served as the Principal Investigator on this project. She is a Founder and Principal of Research for Action. Most recently, her research has focused on the topics of instructional communities, school leadership, organizational learning, and privatization in public education. Another important focus of her work has been on the use of research to inform policy and practice. She has worked extensively with teachers, principals, parents, students and other public school activists to incorporate research and reflection into their efforts to improve urban public schools.

Ruth Curran Neild

Ruth Curran Neild, Ph.D. served as a Co-Principal Investigator on this project. She is a Research Scientist at the Johns Hopkins University. Her scholarly interests, broadly speaking, focus on improving educational outcomes for urban youth through transforming their school experiences. She has published in the areas of high school choice, teacher quality, the ninth grade transition, high school reform, and high school graduation and dropout. She is committed to communicating clearly about research findings to practitioners and policy-makers and is a frequent presenter at conferences and workshops.

Katrina Bulkley

Katrina Bulkley, Ph.D. served as Co-Principal Investigator on this project. She is an Associate Professor of Educational Leadership at Montclair State University. Her work explores the role of governance changes in educational reform. Her recent studies have focused on the role of for-profit and non-profit management organizations in the operations of public schools nationally and in Philadelphia. She is the editor (with Priscilla Wohlstetter) of *Taking Account of Charter Schools: What's Happened and What's Next?* (2004, Teachers College Press) and (with Lance Fusarelli) of "The politics of privatization in education: The 2007 Yearbook of the Politics of Education Association."

Suzanne Blanc

Suzanne (Sukey) Blanc, Ph.D. is an educational anthropologist and a former middle school math teacher. She is a senior research consultant at Research for Action and is the founder of Creative Research and Evaluation Services. Her work centers on program evaluation and participatory research in urban schools and communities. She has conducted numerous evaluations of National Science Foundation projects in science, technology, and engineering and also has a long-standing interest in the connection between education and other aspects of urban life such as community arts, community, revitalization, and community organizing.

Roseann Liu

Roseann Liu is a Ph.D. student at the University of Pennsylvania's Graduate School of Education pursuing a dual degree in anthropology and education. She is interested in the cultural productions of youth in transnational and diasporic communities. Prior to beginning graduate school, she was a Research Associate at Research for Action.

Cecily Mitchell

Cecily Mitchell is especially interested in school-based interventions to improve the educational experiences and outcomes for students who have been marginalized within the educational system. Her undergraduate thesis was based on a participatory research project that examined how student academic engagement is mediated by school rules and norms together with race and gender in a 2nd grade classroom. Prior to coming to RFA, she worked in a school-based behavioral health program to develop effective classroom interventions for students with emotional/behavioral disabilities.

Eva Travers

Eva Travers, Ph.D. is Professor Emeritus at Swarthmore College where she taught urban education and education policy. She is involved in ongoing research by RFA on system-wide school reforms in Philadelphia. She held a number of administrative positions at Swarthmore College, including Director of the Program in Education, and Associate Dean. She has served on a variety of national working groups and task forces looking at issues of teacher preparation and teacher education.



RESEARCH FOR *ACTION*

3701 Chestnut Street
Philadelphia, PA 19104
ph 215.823.2500
fx 215.823.2510
www.researchforaction.org